

CENTRO UNIVERSITÁRIO FEI
RODRIGO NAHMIR RUY

SISTEMAS PROATIVOS DE RECOMENDAÇÃO SENSÍVEIS AO CONTEXTO:
autonomia de Sistemas de Recomendação baseado nas informações provenientes da Internet
das Coisas e Aprendizagem Indutiva

São Bernardo do Campo

2017

RODRIGO NAHMIR RUY

SISTEMAS PROATIVOS DE RECOMENDAÇÃO SENSÍVEIS AO CONTEXTO:
autonomia de Sistemas de Recomendação baseado nas informações provenientes da Internet
das Coisas e Aprendizagem Indutiva

Dissertação apresentado ao Centro
Universitário FEI, como parte dos requisitos
necessários para obtenção do título de Mestre
em Engenharia Elétrica. Orientado pelo Prof.
Dr. Plinio Thomaz Aquino Jr.

São Bernardo do Campo

2017

Nahmir Ruy, Rodrigo.

Sistemas Proativos de Recomendação Sensíveis ao Contexto:
Autonomia de Sistemas de Recomendação baseado nas informações
provenientes da Internet das Coisas e Aprendizagem Indutiva / Rodrigo
Nahmir Ruy. São Bernardo do Campo, 2017.

104 f. : il.

Dissertação - Centro Universitário FEI.

Orientador: Prof. Dr. Plinio Thomaz Aquino Jr..

Coorientador: Prof. Dr. Rodrigo Filev Maia.

1. Internet das Coisas. 2. Sistemas de Recomendação. 3. Sistemas de
Recomendação Sensíveis ao Contexto. 4. Sistemas Proativos de
Recomendação. I. Thomaz Aquino Jr., Plinio , orient. II. Título.

Aluno: Rodrigo Nahmir Ruy

Matrícula: 115117-4

Título do Trabalho: Sistemas proativos de recomendação sensíveis ao contexto: autonomia de sistemas de recomendação baseado nas informações provenientes da internet das coisas e aprendizagem indutiva.

Área de Concentração: Inteligência Artificial Aplicada à Automação

Orientador: Prof. Dr. Plinio Thomaz Aquino Junior

Data da realização da defesa: 20/06/2017

ORIGINAL ASSINADA

Avaliação da Banca Examinadora:

São Bernardo do Campo, 20 / 06 / 2017.

MEMBROS DA BANCA EXAMINADORA

Prof. Dr. Plinio Thomaz Aquino Junior Ass.: _____

Prof. Dr. Marcel Jacques Simonette Ass.: _____

Prof. Dr. Flavio Tonidandel Ass.: _____

A Banca Julgadora acima-assinada atribuiu ao aluno o seguinte resultado:

APROVADO

REPROVADO

VERSÃO FINAL DA DISSERTAÇÃO

**APROVO A VERSÃO FINAL DA DISSERTAÇÃO EM QUE
FORAM INCLUÍDAS AS RECOMENDAÇÕES DA BANCA
EXAMINADORA**

Aprovação do Coordenador do Programa de Pós-graduação

Prof. Dr. Carlos Eduardo Thomaz

Aos meus pais, professores e amigos presentes
na minha vida.

AGRADECIMENTOS

Agradeço aos meus pais, pelo amor, incentivo e apoio incondicional. Apesar de todas as dificuldades eles me fortaleceram e sempre me fizeram entender que o futuro é moldado conforme as decisões que tomamos no presente.

Agradeço aos professores Dr. Plinio Thomaz e Dr. Rodrigo Filev, pelas orientações, apoio e confiança durante a elaboração desta dissertação, onde através de seus conhecimentos me proporcionaram um grande aprendizado e amadurecimento como pesquisador.

Ao Centro Universitário FEI e Telefônica-Vivo por proporcionar as condições necessárias para o desenvolvimento desta pesquisa.

Aos integrantes do laboratório de Inovação e Internet das Coisas, que através de um convívio harmonioso foram importantes para geração de ideias e realização desta pesquisa.

Enfim, agradeço a todos que de forma direta ou indireta fizeram parte do desenvolvimento desta dissertação.

“Não é preciso ter olhos abertos para ver o sol,
nem é preciso ter ouvidos afiados para ouvir o
trovão. Para ser vitorioso você precisa ver o que
não está visível”.

Sun Tzu

RESUMO

A Internet das Coisas (IoT) está cada vez mais presente no cotidiano das pessoas e nos seus pertences. Como consequência desta evolução, grandes quantidades de dados dos mais diversos tipos são originados. Devido a IoT ter como premissa a modelagem virtual de coisas do mundo físico, estes dados podem ser tratados e utilizados para diferentes propósitos. Quando Sistemas de Recomendação tradicionais utilizam destes dados provenientes da IoT para caracterizar o contexto onde estão inseridos, surge o conceito de Sistemas de Recomendação Sensíveis ao Contexto. Embora grande parte dos sistemas sensíveis ao contexto utilizem destes dados apenas para a melhoria de suas sugestões, dados do contexto podem determinar também situações onde o sistema deva fazer uma recomendação proativa (autônoma), ou seja, o sistema fornece uma recomendação sem que o usuário peça explicitamente por uma sugestão. Uma vez que o Sistema de Recomendação tenha autonomia para fazer uma sugestão, surge o conceito de Sistemas Proativos de Recomendação. Atualmente poucas pesquisas aplicam o aspecto de proatividade para gerar recomendações, portanto existem grandes lacunas sobre as características que devem ser levadas em consideração ao se desenvolver estes sistemas e como estas características se interrelacionam. O principal problema abordado neste trabalho é a identificação de quais são as dificuldades e aspectos que são inerentes ao desenvolvimento de Sistemas Proativos de Recomendação. Para tanto, artigos envolvendo os conceitos de Internet das Coisas, Sistemas de Recomendação, Sistemas de Recomendação Sensíveis ao Contexto, Sistemas Proativos de Recomendação e métodos de Aprendizado de Máquina, são debatidos e servem como base teórica da pesquisa. Uma vez identificado os principais aspectos e dificuldades, foi desenvolvido um modelo de avaliação denominado Sistemas Proativos de Recomendação Sensíveis ao Contexto, onde este é dividido em 4 pilares. Para aplicação deste modelo de avaliação, um Sistema Proativo de Recomendação de filmes, que utiliza dados simulados da IoT e método de Aprendizado de Máquina Indutivo, foi elaborado. Como resultado, o modelo de avaliação foi desenvolvido contemplando os aspectos mais importantes que caracterizam os Sistemas Proativos de Recomendação. Além disto foi desenvolvido um Sistema Proativo de Recomendação Sensível ao Contexto para sugestões proativas de filmes. O algoritmo denominado C4.5 foi utilizado para a geração de árvores de decisão que determinam modelos de preferências de filmes e para gerar árvores de decisão que determinam momentos propícios, para sugestões proativas de filmes aos seus usuários.

Palavras-chave: Internet das Coisas. Sistemas de Recomendação. Sistemas de Recomendação Sensíveis ao Contexto. Sistemas Proativos de Recomendação.

ABSTRACT

More and more Internet of Things (IoT) is present in human being diary life and their belongings. Because of this evolution, large quantities of data of different types are generated. Due to the IoT has as premise the virtual modeling of things in physical world, these data can be treated and used for different purposes. When Recommender Systems used data originating from IoT to characterize the context where are inserted, the concept of Recommender Systems Context-awareness arise. Although a great part of systems context-awareness use these data for increase the quality of suggestions, context data can also determinate situations where the system must make a proactive recommendation, in other words, the system provides a recommendation without explicit request of user. Once the Recommender System has autonomy to provide suggestions, the concept of Proactive Recommender System arise. Nowadays few researches applied the aspect of proactivity to provide recommendations, thus there are large gaps about the characteristics that must be considerate when develop these systems and how these characteristics are interrelated. The main problem addressed in this dissertation it is the identification of which difficulties and aspects inherent in the development of Proactive Recommender Systems. For this purpose, articles involving studies about Internet of Things, Recommender Systems, Recommender Systems Context-awareness, Proactive Recommender Systems and Machine Learning methods, are debated and serve as theoretical basis of this research. Once the principal aspects and difficulties are identified, was developed an evaluation model called Proactive Recommender System Context-Awareness, where this model is divide into four parts. For application of this evaluation model, a Proactive Recommender System of movies, that use simulated IoT data and Inductive Machine Learning, was elaborated. As a result, the evaluation model was deployed with the principal aspects that characterize the Proactive Recommender System. Besides that, was developed a Proactive Recommender System Context-Awareness of movies. The algorithm C4.5 was used for decision trees generation that provide users' preferences profile of movies and to determinate propitious moments for proactive recommendations of movies.

Keywords: Internet of Things. Recommender System. Recommender System Context-Awareness. Proactive Recommender System.

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo de Sistemas Proativos de Recomendação Mobile	18
Figura 2 - Modelo de Salman et al. (2015) para SPR.....	20
Figura 3 - Tecnologias de informações, comunicações e IoT	29
Figura 4 - Visão técnica geral da IoT	30
Figura 5 - Projeções de Dispositivos Conectados por Pessoa	32
Figura 6 - Porcentagens das áreas em IoT	34
Figura 7 - Contexto Primário e Secundário	36
Figura 8 - Ciclo de Vida das análises de Contexto.....	41
Figura 9 - Estrutura Contexto em IoT do modelo de avaliação de SPRSC.....	45
Quadro 1 - Vantagens e Desvantagens SR Baseado em Conteúdo	48
Quadro 2 – Aspectos para desenvolvimento de SR com Filtragem Colaborativa.....	50
Quadro 3 - Vantagens e Desvantagens SR Filtragem Colaborativa.....	52
Quadro 4 - Vantagens e Desvantagens SR Demográficos	53
Quadro 5 - Vantagens e Desvantagens SR Funções de Utilidade	55
Quadro 6 - Vantagens e Desvantagens SR Baseado em Conhecimento	57
Quadro 7 - Características de técnicas individuais em Sistemas de Recomendação.....	57
Figura 10 - Estrutura SR de SPRSC	62
Figura 11 - Estrutura Usuário de SPRSC	65
Figura 12 - Métodos de Aprendizado de Máquina Utilizados.....	71
Figura 13 - Tempo vs Erro de técnicas de Aprendizado de Máquina em SR na IoT	71
Figura 14 - Pseudocódigo C4.5	74
Figura 15 - Base de Treinamento com Valores Contínuos.....	76
Figura 16 - Estrutura Aprendizado de Máquina em SPRSC	78
Figura 17 – Modelo de Avaliação de SPRSC	80
Figura 18 - Acertos de acordo com variação de Instâncias	85
Figura 19 - Árvore de Decisão Podada.....	85
Quadro 8 – Estrutura Usuários do Modelo de Avaliação	89
Quadro 9 - Estrutura Contexto em Internet das Coisas do Modelo de Avaliação.....	90
Quadro 10 - Estrutura Aprendizado de Máquina do Modelo de Avaliação	91
Quadro 11 - Estrutura Sistemas de Recomendação do Modelo de Avaliação	91

SUMÁRIO

1	INTRODUÇÃO	13
1.1	OBJETIVO.....	14
1.2	MOTIVAÇÃO	14
1.3	ORGANIZAÇÃO DO TRABALHO	14
2	DEFINIÇÕES DE SISTEMAS DE RECOMENDAÇÃO E SUAS VARIAÇÕES	16
2.1	DEFINIÇÃO DE SISTEMA DE RECOMENDAÇÃO	16
2.2	DEFINIÇÃO DE SISTEMAS DE RECOMENDAÇÃO SENSÍVEIS AO CONTEXTO	17
2.3	DEFINIÇÃO E ASPECTOS DE SISTEMAS PROATIVOS DE RECOMENDAÇÃO.....	17
2.4	SISTEMAS PROATIVOS DE RECOMENDAÇÃO SENSÍVEIS AO CONTEXTO.....	21
3	CONTEXTO EM INTERNET DAS COISAS	27
3.1	ORIGENS E DEFINIÇÕES DA IOT.....	27
3.2	PERSPECTIVAS E GRUPOS DE APLICAÇÃO	32
3.3	CONTEXTO	34
3.4	CONTEXTO EM IOT.....	38
3.4.1	Redes de sensores em IoT	38
3.4.2	Sensibilidade ao contexto através da IoT e seus aspectos	39
3.5	ESTRUTURA CONTEXTO EM IOT DE SPRSC	44
4	SISTEMAS DE RECOMENDAÇÃO	46
4.1	SR BASEADOS EM CONTEÚDO	46
4.2	SR COM FILTRAGEM COLABORATIVA.....	49
4.3	SR DEMOGRÁFICOS.....	53
4.4	SR FUNÇÕES DE UTILIDADE	54
4.5	SR BASEADO EM CONHECIMENTO	56
4.6	QUADRO DEMONSTRATIVO ENTRE OS MÉTODOS DE SR	57
4.7	SR HÍBRIDOS	58
5	USUÁRIOS	63
5.1	USUÁRIOS.....	63
5.2	ESTRUTURA USUÁRIOS DE SPRSC	64
6	APRENDIZADO DE MÁQUINA	66
6.1	INTELIGÊNCIA ARTIFICIAL.....	66

6.2	APRENDIZADO DE MÁQUINA	66
6.3	APRENDIZAGEM INDUTIVA	69
6.4	APRENDIZAGEM DE MÁQUINA EM SPR	70
6.5	APRENDIZAGEM INDUTIVA ATRAVÉS DE ÁRVORES DE DECISÃO	71
6.6	ESTRUTURA APRENDIZADO DE MÁQUINA DE SPRSC	77
7	MODELO DE AVALIAÇÃO ESTRUTURADO DE SISTEMAS PROATIVOS DE RECOMENDAÇÃO SENSÍVEIS AO CONTEXTO E SIMULAÇÃO	79
7.1	BASES DE DADOS DO SPRSC DE FILMES	79
7.1.1	Base de dados sala IoT	79
7.1.2	Base de dados avaliações de filmes	82
7.2	APRENDIZADO DE MÁQUINA	83
7.2.1	C4.5 aplicado na base de dados sala IoT	83
7.2.2	C4.5 aplicado na base de dados avaliações de filmes	85
7.3	VALOR DE LIMIAR.....	86
7.4	TÉCNICA DE SISTEMA DE RECOMENDAÇÃO	89
7.5	AVALIAÇÃO DO SPRSC E CONSIDERAÇÕES FINAIS DA SEÇÃO.....	89
8	CONCLUSÃO E TRABALHOS FUTUROS.....	97
8.1	CONCLUSÃO	97
8.2	TRABALHOS FUTUROS.....	98
	REFERÊNCIAS	100

1 INTRODUÇÃO

O mundo não está apenas crescendo, está crescendo de forma vertiginosa, não em suas propriedades geológicas, mas sim em termos tecnológicos e populacionais. Estima-se que no ano de 2020 existirá uma população mundial de 7,6 bilhões de pessoas e 50 bilhões de dispositivos conectados com alguma rede de comunicação. O quociente da divisão destes dispositivos pela população mundial resultará em uma média de 6,58 dispositivos por pessoa (AL-FUQAHA et al., 2015).

Internet das Coisas (IoT) tem como premissas a comunicação de qualquer coisa, a qualquer momento e em qualquer lugar (INTERNATIONAL TELECOMMUNICATION UNION - ITU-T, 2012), portanto os dados gerados por diversos dispositivos podem ser tratados por diferentes sistemas de acordo com suas finalidades. Outra característica importante da IoT, é que as coisas presentes no mundo se conectem com a Internet e comuniquem-se umas com as outras com o mínimo de intervenção dos seres humanos. Seu objetivo principal é criar um mundo melhor para a humanidade, onde as coisas saibam o que os seres humanos gostam, desejam, precisam e hajam de forma coerente sem instruções explícitas (PERERA et al., 2014).

Das premissas e objetivos citados acima, determina-se algumas características que devem ser compreendidas pelos sistemas inseridos dentro do âmbito da IoT. Quando se diz mínimo de intervenção humana e que o sistema haja de forma coerente sem instruções explícitas, se faz a necessidade do desenvolvimento de sistemas proativos. Ao dizer que as coisas saibam o que os seres humanos gostam, desejam e precisam, fica claro que os sistemas devem também conhecer as preferências de seus usuários. Dado tais fatos, algumas questões devem ser previamente pensadas, tais como: de que forma os dados provenientes da IoT possam ser consumidos por seus sistemas afim de que estes sistemas procedam de forma proativa. Outra questão é como estes sistemas devem ser desenvolvidos para atenderem os gostos, anseios e necessidades de seus usuários. Estas questões, quando tratadas de forma individuais, apresentam conceitos que auxiliam nas resoluções, porém quando tratadas de forma conjunta surgem grandes desafios e muitas dúvidas.

Com o intuito de colaborar para a solução das questões inerentes aos sistemas desenvolvidos no âmbito da IoT, neste trabalho foi desenvolvido um modelo de avaliação para Sistemas Proativos de Recomendação Sensíveis ao Contexto. Este modelo é dividido em quatro estruturas, sendo elas: Usuários, Contexto em Internet das Coisas, Aprendizado de Máquina e Sistemas de Recomendação. Para cada estrutura, ao longo da dissertação, é discutido suas características e como estas podem ser empregadas em ambientes IoT.

1.1 OBJETIVO

O objetivo deste trabalho é elaborar um modelo de avaliação de Sistemas de Recomendação Sensíveis ao Contexto que permita aos interessados ter uma análise mais aprofundada do tema e conseqüentemente um aprendizado mais organizado e minucioso. Para isto, conceitos e pesquisas sobre informações provenientes do contexto através da Internet das Coisas, Sistemas de Recomendação usuais e Aprendizado de Máquina foram analisados, explicados, debatidos e relacionados. Após definido e elaborado o modelo, é feita uma simulação de um Sistema Proativo de Recomendação Sensível ao Contexto que faz sugestões de filmes de forma autônoma (proativa). Para o desenvolvimento deste sistema, foi aplicado um método de aprendizado de máquina indutivo de árvores de decisão em dados provenientes de uma simulação teórica de uma sala de Internet das Coisas e aplicado também em dados referentes a diversas avaliações de filmes obtidas de diversos usuários. Posteriormente a simulação do sistema, este é avaliado e debatido conforme o modelo desenvolvido.

1.2 MOTIVAÇÃO

Dado que um dos grandes temas discutidos atualmente no âmbito tecnológico é a Internet das Coisas e devido este ser um conceito relativamente novo, grandes desafios relacionados a questões estruturais e aplicações necessitam de pesquisa.

Além dos desafios, outra motivação para a escolha deste tema foi identificar como as características da IoT são utilizadas para se determinar diferentes contextos e como estes contextos podem ser usados para permitir que sugestões sejam feitas de forma autônoma (proativa).

1.3 ORGANIZAÇÃO DO TRABALHO

Esta dissertação compreende as seguintes seções:

- a) a seção 2 descreve as definições de Sistemas de Recomendações e suas variações.

Nesta seção é apresentado a base teórica e as estruturas fundamentais que definem um Sistema Proativo de Recomendação Sensível ao Contexto;

- b) na seção 3 é apresentado algumas definições de Internet das Coisas e Contexto, além do que, é apresentado como a Internet das Coisas é utilizada para a caracterização de

diferentes contextos e como ela pode ser inserida nos diversos Sistemas de Recomendação;

- c) na seção 4 estão explicadas as técnicas mais tradicionais de Sistemas de Recomendação;
- d) a seção 5 explica o que é um usuário de um Sistema de Recomendação e como ele pode ser classificado;
- e) a seção 6 apresenta as características da estrutura Aprendizado de Máquina do modelo de avaliação de Sistemas Proativos de Recomendação Sensíveis ao Contexto. Como o Sistema Proativo de Recomendação de filmes simulado ao final desta dissertação utiliza do algoritmo C4.5, nesta seção também são explicados o funcionamento e as características deste algoritmo;
- f) na seção 7 está o modelo de avaliação de Sistemas Proativos de Recomendação Sensível ao Contexto completo. Nesta seção também é simulado e avaliado conforme o modelo, um Sistema Proativo de Recomendação que faz sugestões de filmes;
- g) na seção 8 está a conclusão da dissertação e trabalhos futuros.

2 DEFINIÇÕES DE SISTEMAS DE RECOMENDAÇÃO E SUAS VARIAÇÕES

Nesta seção são apresentadas definições para Sistemas de Recomendação (SR), Sistemas de Recomendação Sensíveis ao Contexto (SRC), Sistemas Proativos de Recomendação (SPR) e alguns artigos selecionados que auxiliam na determinação de um modelo de avaliação de SPR denominado Sistemas Proativos de Recomendação Sensíveis ao Contexto (SPRSC).

2.1 DEFINIÇÃO DE SISTEMA DE RECOMENDAÇÃO

Ao buscar um novo item, ação e/ou serviço para entretenimento ou trabalho, o método mais convencional e antigo é o compartilhamento presencial de opiniões entre pessoas. Devido a quantidade de horas disponíveis para a troca de informações presenciais nos dias de hoje serem escassas e também existirem inúmeros itens, ações e/ou serviços, muitas vezes a troca de informações e sugestões provenientes destas conversas não satisfazem as dúvidas dos usuários (BOBADILLA et al., 2013; SCHAFER et al., 2007).

Com o intuito de economizar o tempo gasto em pesquisas e ajudar no aprimoramento da qualidade das sugestões, ferramentas automáticas foram desenvolvidas resultando nos Sistemas de Recomendação (SR). Estes sistemas adquirem e armazenam informações sobre as preferências de seus usuários, sendo que estas informações podem ser adquiridas de forma explícita, através da coleta das avaliações provenientes dos usuários, e de forma implícita, através da determinação de como o usuário se comporta com os itens, ações e/ou serviços inseridos nos SR (BOBADILLA et al., 2013).

Conforme Burke (2002), todos os SR estão fundamentados em três divisões estruturais, sendo elas:

- a) dados preliminares: contêm as informações já conhecidas pelos sistemas antes que o processo de recomendação se inicie;
- b) dados inseridos: contempla as informações inseridas pelos usuários que posteriormente são utilizadas no processo de elaboração das recomendações; e
- c) processo: aplica algoritmos de aprendizagem, para combinar os dados preliminares com os dados inseridos pelos usuários e gerar recomendações.

Em busca do aprimoramento da qualidade e utilidade das sugestões geradas pelos SR, novos conceitos emergentes e em evolução nos dias de hoje, tais como Computação Ubíqua e

Internet das Coisas (*Internet of Things* – IoT), são integrados nos SR. Através desta integração, informações obtidas do contexto onde estão inseridos os SR são utilizados para a elaboração de sugestões (PERERA et al., 2014; SALMAN et al., 2015).

2.2 DEFINIÇÃO DE SISTEMAS DE RECOMENDAÇÃO SENSÍVEIS AO CONTEXTO

Sistemas de Recomendação Sensíveis ao Contexto são caracterizados quando dados provenientes do contexto onde seus usuários estão inseridos são levados em consideração para a geração das sugestões de itens, serviços e/ou ações. Alguns exemplos destes dados são: interações de um usuário com um recurso qualquer, horário, localização, temperatura, estação do ano ou qualquer dado que possa ser determinado do Contexto (MUNOZ-ORGANERO et al., 2010; SALMAN et al., 2015).

Em outras palavras, Adomavicius e Tuzhilin (2015) definiram Sistemas de Recomendação Sensíveis ao Contexto como sistemas que objetivam a modelação e predição de gostos e preferências de seus usuários, levando em conta dados do Contexto onde estão inseridos o usuário e sistema. Estes autores utilizam de 3 domínios fundamentais para caracterizar um Sistema de Recomendação Sensível ao Contexto, sendo eles: Usuários, sendo que este domínio engloba os usuários para quem são feitas as recomendações. Itens, engloba os itens, ações e/ou serviços que os sistemas podem fornecer aos seus usuários e por fim o domínio Contexto, onde estão inseridos todos os dados que caracterizam o contexto onde estão os usuários e os sistemas de recomendação.

2.3 DEFINIÇÃO E ASPECTOS DE SISTEMAS PROATIVOS DE RECOMENDAÇÃO

SR tradicionais, são incumbidos de determinar quais itens, ações e/ou serviços recomendar, sendo que as recomendações só devem ser feitas quando solicitadas de forma explícita pelos usuários. Se contrapondo a esta característica, Sistemas Proativos de Recomendação (SPR) além de determinar quais itens, ações e/ou serviços sugerir, estes devem também determinar quais momentos são propícios para sugestões (WOERN DL et al., 2011).

Devido a autonomia dos SPR, estes sistemas devem assegurar que suas sugestões sejam fornecidas no momento correto, no local correto, de uma forma correta e para a pessoa certa. Caso tais características não sejam atendidas, ocorrerá conflitos e descontentamento entre o usuário e sistema. Em suma, SPR tem como pretensão antecipar as necessidades de seus

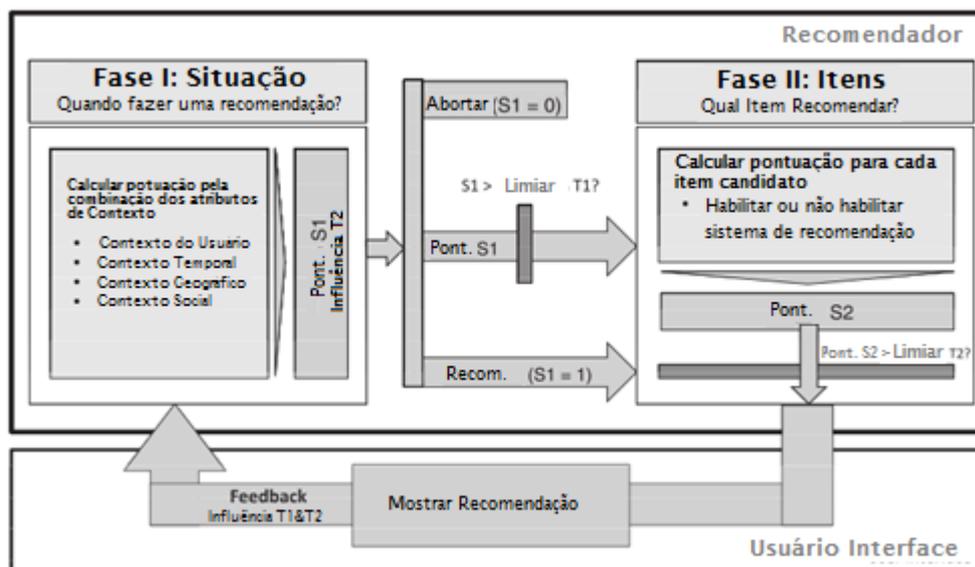
usuários e recomendar itens, ações e/ou serviços apropriados (BRAUNHOFER et al., 2015; SABIC, 2016).

A determinação de quais são os momentos e locais ideais para sugestões autônomas, ainda apresentam poucas pesquisas relacionadas, portanto os SPR apresentam incertezas em relação ao seu desenvolvimento e aplicabilidade (SABIC, 2016).

Com intuito de propor um modelo de como um SPR deve ser desenvolvido, Bedi e Agarwal (2012), propõe que a elaboração destes sistemas deve ser dividida em duas fases características. A primeira é a avaliação da situação (determinação do contexto) e a segunda é a avaliação de recomendações, sendo que a segunda fase só é iniciada quando a primeira apresenta uma situação propícia para a sugestão.

Gallego, Woerndl e Huecas (2013), define um SPR também em duas fases. Conforme figura 1, a primeira fase consiste em obter informações do contexto onde o sistema e seus usuários estão inseridos. A segunda fase é responsável por determinar quais itens devem ser recomendados. Para cada fase é atribuído um valor. O valor $S1$ caracteriza o nível de pertinência para uma sugestão autônoma da primeira fase, ou seja, dado um contexto em específico é atribuído um valor que quantifica a qualidade do momento para uma sugestão proativa. O valor $S2$ caracteriza o nível de avaliação de um item que pode ser recomendado na segunda fase. Ambas as fases possuem valores de limiar, ou seja, valores que devem ser superados para o SPR prosseguir com uma sugestão proativa. O valor $T1$ representa o valor de limiar para a primeira fase e $T2$ para a segunda fase.

Figura 1 – Modelo de Sistemas Proativos de Recomendação Mobile



Fonte: Autor "adaptado de" Gallego, Woerndl e Huecas, 2013

Concluindo, tal modelo apresenta uma característica importante que deve ser levado em consideração ao se desenvolver SPR, sendo ela: A fase de determinação do contexto deve influenciar no nível de urgência dos itens a serem recomendados na fase de sugestão. Tal característica permite apurar e valorar as necessidades que seus usuários têm ao longo do tempo.

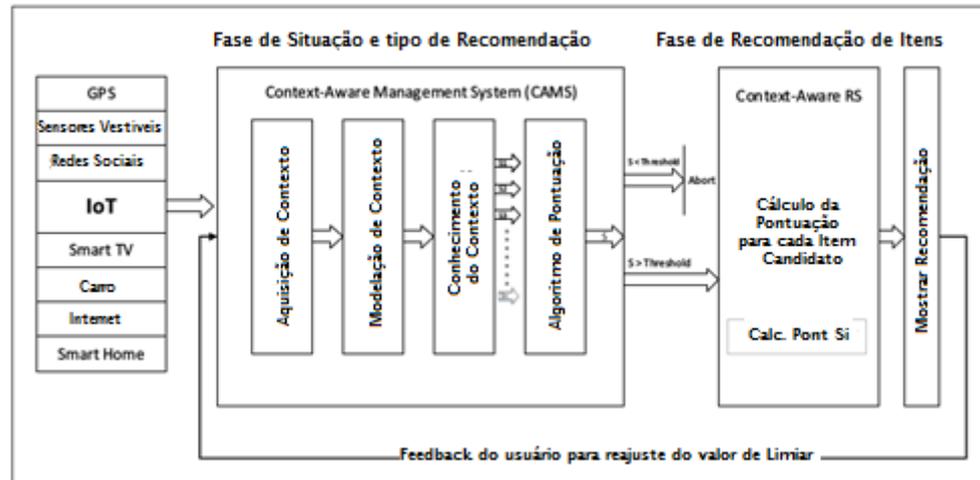
Prosseguindo com pesquisas relacionadas aos SPR, Salman et al. (2015) em seu artigo exemplifica três situações que colaboram para determinar algumas características inerentes a estes sistemas, sendo eles:

- a) suponha que um veículo possua sensores que permitam aferir o nível de combustível e sua localização seja conhecida devido ao GPS. O SPR, verifica a quantidade de combustível e caso este note que a quantidade está na reserva, o sistema faz uma varredura dos postos de combustíveis próximos e faz sugestões daqueles que apresentam boas avaliações;
- b) sistemas desenvolvidos para restaurantes, além de levar em consideração localização e o histórico de comidas prediletas de seus clientes, os mesmos devem ser capazes de diferenciar qual tipo de comida sugerir. Como exemplo, de acordo com o horário o mesmo deve fazer sugestões levando em consideração se é café da manhã, almoço ou jantar; e
- c) submergindo tais sistemas em um cenário médico e de saúde pessoal que compreende os conceitos da IoT, um usuário em particular que sofra com pressão alta e/ou diabetes está sendo frequentemente acompanhado por sensores de medições capazes de aferir a pressão e indicar os níveis de açúcar no sangue. Ao menor sinal de alguma alteração em tais níveis e que caracterize uma emergência, o sistema deve ser capaz de identificar tal urgência e recomendar hospitais e/ou farmácias próximos a ele.

Tais cenários ajudam a caracterizar níveis de situações emergenciais. Observe que a primeira e terceira situação supracitadas apresentam níveis de urgências maiores comparados com a segunda. SPR devem ser capazes de distinguir entre situações emergenciais e corriqueiras, de tal forma que o mesmo tome decisões de forma particular para cada situação (SALMAN et al., 2015).

Em seu artigo Salman et al. (2015), propõem um SPR para um ambiente inserido na IoT. Conforme figura 2, este sistema é dividido em três blocos, sendo eles: IoT, Contexto e SR.

Figura 2 - Modelo de Salman et al. (2015) para SPR



Fonte: Autor “adaptado de” Salman et al., 2015

Salman et al. (2015) explica a funcionalidade que cada bloco exerce dentro deste sistema, sendo elas:

- a) o bloco referente a IoT, compreende o conjunto de coisas inseridos na IoT. Estas coisas contribuem no envio de dados referente ao contexto em que está inserido o sistema;
- b) o segundo bloco denominado Fase de Situação e Tipo de Recomendação, tem como finalidade fazer as tratativas referente as informações obtidas do contexto e determinar quais são os momentos corretos para fazer sugestões. Note que o mesmo é subdividido em quatro estágios, sendo eles:
 - aquisição de contexto: Neste estágio são armazenados os dados brutos provenientes dos elementos submergidos na IoT;
 - modelação de contexto: Em tal etapa, os dados armazenados no estágio anterior, são processados e servirão como base de informações para aprendizagem do contexto;
 - conhecimento do contexto: Através dos dados já tratadas provenientes do contexto, estas informações são relacionadas e mostram a situação em que se encontra o sistema em relação ao contexto; e
 - algoritmo de pontuação: Por fim, a partir das informações do contexto proveniente da etapa anterior, é calculado um valor (através de métodos de aprendizado de máquina) que determina se a situação é pertinente ou não para uma futura recomendação proativa. Note que este valor calculado é comparado a um valor de limiar pré-determinado pelo sistema (com base em situações já

acontecidas) e se o mesmo for superior o sistema entra na próxima fase, caso contrário o sistema não faz recomendação alguma.

- c) o terceiro e último bloco de tal modelo é denominado Fase de Recomendação de Itens. Como dito anteriormente, este só é executado quando habilitado pelo bloco anterior. Em tal bloco, estão inseridas as técnicas que são utilizadas pelos SR para fornecimento de sugestões aos seus usuários.

Em suma, Sistemas Proativos de Recomendação originaram-se dos Sistemas de Recomendação tradicionais e Sistemas de Recomendação Sensíveis ao Contexto. A grande diferença dos sistemas proativos em relação aos outros, está em determinar momentos propícios para uma sugestão do sistema de forma autônoma. Para isso, valores que relacionam o quão um item, ação e/ou serviço é do gosto de um usuário e o quão um momento é propício para uma sugestão, são relacionados e determinam se o sistema deve fornecer uma recomendação autônoma. Esta característica é de fundamental importância ao desenvolver um modelo estrutural para Sistemas Proativos de Recomendação Sensíveis ao Contexto.

2.4 SISTEMAS PROATIVOS DE RECOMENDAÇÃO SENSÍVEIS AO CONTEXTO

Um dos objetivos desta dissertação é definir quais são os parâmetros e como estes se relacionam afim de determinar momentos pertinentes para sugestão em Sistemas Proativos de Recomendação. Para tanto, uma série de artigos relacionados diretamente e indiretamente aos SPR foram revisados, destacando-se os explanados nesta seção de embasamento teórico.

Bedi e Agarwal (2012), propõe um modelo teórico para um SPR para sugestões de restaurantes. Para tanto, a arquitetura de seu sistema é dividida em duas fases, sendo a primeira responsável pela determinação do contexto correto para fazer uma recomendação e a segunda qual recomendação fazer. Em relação a primeira fase, este utiliza quatro conjuntos de determinação de contexto, sendo eles: Distância, Tempo, Preço e Acessibilidade. Tais características são tratadas com regras em lógica Fuzzy. Após habilitada a segunda fase, é utilizada de uma abordagem colaborativa para determinação de quais recomendações o sistema deve fazer. Tal sistema não foi implementado de forma prática, porém uma pesquisa foi realizada com pessoas para verificar a pertinência e o grau de aceitabilidade deste sistema. Através dos resultados, foi comprovado a eficácia e aceitação deste SPR nesta situação em específico.

Ainda sobre SPR, Gallego et al. (2013) apresenta um estudo para mostrar a importância

dos SPR porém em ambientes *e-Learning*. Para tanto, foram realizadas pesquisas com 104 pessoas, sendo elas professores ou cientistas. A pesquisa baseava-se em determinar quais os níveis de importância e pertinência de acordo com a situação (contexto) que faziam tais usuários utilizarem da ferramenta *e-Learning*. Para caracterização do contexto os seguintes conjuntos foram contemplados: Geográfico, mostra se o usuário está ou não no seu espaço de trabalho; Período do dia, descreve se é manhã, tarde, final da tarde ou noite; Atividade, responsável por caracterizar qual tarefa o usuário está executando dentro do ambiente *e-Learning*. Após realizado tais pesquisas e estas dispostas em gráficos, conclui-se que os dados provenientes do contexto influenciam diretamente nos momentos em que os usuários necessitam de recomendações ou não, justificando uma visão de autonomia em SR. Destaca-se neste artigo que através de cálculos matemáticos são determinados valores numéricos que servem para determinação se a situação é pertinente ou não para uma sugestão proativa.

Gallego, Woerndl e Huecas (2013), desenvolveram de forma prática um SPR. Através de uma interface com sistema Android, este SPR apresenta aos seus usuários sugestões de forma proativa dependendo do contexto. Este é dividido em duas fases características, sendo elas: avaliação da situação (contexto) e avaliação da recomendação. Vale destacar que os usuários de tais sistemas podem fornecer *feedbacks* no final de cada fase, ou seja, em relação a primeira fase caso o sistema determine que o contexto é um momento propício e o usuário não queira receber tal sugestão o mesmo pode ativar uma função “agora não” e na segunda fase é possível fazer uma avaliação positiva ou negativa do item sugerido no final da fase de recomendação. Para que o sistema crie regras de avaliação do contexto, foi realizado uma série de perguntas a 58 pessoas. Tais perguntas tinham como objetivo determinar em quais situações uma recomendação autônoma teria valor para os usuários. Vale ressaltar que as recomendações mostradas na interface foram testadas em duas formas, sendo elas: através de *widgets* e através de notificação na barra superior da interface com sistema Android. Dois pontos deste trabalho merecem uma atenção especial, o primeiro, como dito anteriormente e explicado neste capítulo, é referente a interação entre as duas fases de avaliação, onde a primeira interfere nos valores de limiar da segunda. A segunda remete a uma característica importante na avaliação do contexto, que diz que usuários atarefados e com pressa para realização de suas tarefas tendem a não dar atenção a recomendações vindas de sistemas proativos.

Um modelo para SPR que compreende outras características comparados aos artigos tratados anteriormente foi elaborado por Bedi et al. (2014). Denominado SAPRS, tal modelo apresenta como característica uma explicação sobre a relevância e do porquê um item em específico está sendo sugerido. Tal explicação é mostrada juntamente com os itens que foram

recomendados. Assim como os SPR anteriores o sistema é dividido em duas fases: avaliação do contexto e avaliação de qual recomendação fazer. Na primeira fase é utilizado lógica Fuzzy para determinar qual momento é oportuno para realização de uma recomendação e na segunda são utilizadas de técnicas de recomendações denominadas: filtragem colaborativa, LRCF (uma variação da técnica filtragem colaborativa) e pôr fim a mesma técnica LRCF mais a explicação dos itens apresentados para recomendação. Para validar a eficácia de cada técnica apresentada, foram utilizadas de métricas de avaliação denominadas *Precision*, *Recall* e F1. A técnica que apresentou melhor resultado foi LRCF mais explicação, demonstrando que para este contexto o sistema SAPRS obtêm êxito em seus objetivos.

Yao et al. (2014) propôs um Sistema de Recomendação com filtragem colaborativa baseado em abordagens probabilísticas e dados do contexto provenientes da IoT. As seguintes informações são obtidas a partir destes dados: relações entre os usuários, correlação entre os itens e interação entre usuário e item. Métodos de grafos, *cosine similarity*, fatoração de matrizes e fórmulas algébricas foram utilizados pelo sistema para inferir conhecimento a partir destas informações. A IoT neste sistema tem como finalidade a representação de ambientes e objetos físicos em ambientes e objetos virtuais. Para a realização dos experimentos práticos desta pesquisa, diversos ambientes (e.g., cozinha, quarto) foram sensoriadas e seus objetos e/ou itens (e.g., impressoras, computadores, mochilas, TV, cafeteiras, estojo de primeiros socorros) etiquetados com RFID. Através destas modelações, a quantidade e frequência com que estes ambientes e objetos são utilizados foram determinados. Esta pesquisa apresenta uma abordagem probabilística para o tratamento de problemas de *cold start* (tal problema é explicado no capítulo 4 desta dissertação e tratado no artigo de Burke publicado no ano de 2002). Destaca-se que neste artigo é elucidado sobre homofilia, ou seja, pessoas similares podem possuir interesses congêneres, onde este conceito pode ser usado também como ferramenta para a geração de recomendações.

Um modelo para um SPR em três âmbitos de aplicação é proposto por Salman et al. (2015), Os três âmbitos de aplicação são: postos de combustíveis, restaurantes e atrações turísticas. Para tanto, informações do contexto são adquiridos de acordo com os elementos submergidos na IoT. Tais sistemas primam por identificar um momento adequado, plausível e que atenda às necessidades de seus usuários para fazer sugestões. Os seguintes elementos submergidos na IoT são contemplados pelo sistema: GPS, horário, sensor de quantidade de combustível, sensor para verificar se o carro está ligado ou desligado, métodos para aferir se o usuário está ou não com fome e se o dia em questão é feriado. 120 gravações de dados aleatórios referente aos estados dos elementos submergidos na IoT foram gerados originando uma base

de treinamento. Para que tal SPR aprendesse as situações descritas nesta base, foi utilizado do método de aprendizado de máquina Redes Neurais. Uma interface foi desenvolvida para que valores dos elementos pudessem ser configurados por usuários. Assim que inseridos os estados do contexto, o SPR analisava e fazia uma recomendação pertinente conforme o contexto exigia.

Mashal, Alsaryrah e Chung (2016), demonstram técnicas de elaboração de sugestões aplicadas para determinação de quais objetos e serviços oferecer aos seus usuários. Tal artigo tem como pretensão introduzir um modelo formal para a IoT, aplicar algoritmos tradicionais para desenvolvimento de sugestões de serviços da IoT e empregar métricas para avaliações da qualidade das recomendações geradas. Para realização dos experimentos práticos, foi obtido um vasto banco de dados. A partir de empresas que fornecem serviços para IoT, sendo elas Libelium, Telus e Blue Rover, tal banco foi desenvolvido compreendendo 110 objetos e 90 serviços. Em relação aos usuários deste SR, 400 indivíduos foram selecionados. Para cada usuário foi perguntado quais objetos tinham ou gostariam de ter, além disso, foi pedido que selecionassem os serviços que eram de seus interesses. Todos os usuários foram indagados de forma semelhante e os mesmos foram separados em dois conjuntos: 80% foram utilizados para a criação de um banco de dados de treinamento, sendo que através destes dados o sistema aprende quais recomendações sugerir, e o 20% restante para verificar a qualidade das recomendações. Dentre os algoritmos utilizados, os que tiveram maior eficácia e qualidade nas recomendações foram as abordagens híbridas, ou seja, aqueles que faziam a integração entre diversos algoritmos.

Na pesquisa de Sachdeva, Dhir e Kumar (2016), foram aplicados diversos métodos de aprendizado de máquina em uma base de dados de restaurantes em Chicago com o intuito de compreender as características que cada restaurante possui e posteriormente fazer recomendações de forma autônoma aos usuários dos SPR. Dentre as características pertencentes a cada restaurante destacam-se: preço, qualidade do ambiente, localização, tipos de comida servidas e avaliações. As técnicas de aprendizado de máquina utilizadas foram: *GA-ANN*, *ANN*, *AdaBoost*, *Decision Tree* e *Bagging*. Três métodos foram utilizados para avaliação de cada forma de aprendizado, sendo elas: Acurácia, que consiste em fazer testes em exemplos não utilizados na base de treinamento. Precisão de erro, afere a variação de um resultado obtido comparado com um resultado que era esperado, e por fim o tempo que cada método de aprendizado demora em seu aprendizado. Destaca-se que os algoritmos de aprendizado *GA-ANN* e *Decision Tree* apresentaram ótimos resultados nos três parâmetros para avaliação. Apesar de não ser utilizada diretamente, a IoT é citada como ferramenta de auxílio de sensoriamento dos usuários para determinação de suas necessidades e gostos.

Todos os artigos descritos acima, auxiliaram na obtenção de quais são os principais parâmetros que devem ser levados em consideração para se avaliar diferentes SR que levam em consideração o Contexto para elaboração de sugestões. Inicialmente, quatro conceitos fundamentais foram obtidos através da análise dos diversos artigos citados acima, sendo eles:

- a) contexto em IoT: nos artigos de Yao et al. (2014) e Salman et al. (2015), foram desenvolvidos SR que utilizam de dados e/ou informações do contexto provenientes da IoT para elaboração de suas sugestões. Estes dados e/ou informações quando adquiridos através das coisas da IoT, auxiliam na determinação do contexto onde os SR estão inseridos. Os artigos de Bedi e Agarwal (2012), Gallego et al. (2013), Gallego, Woerndl e Huecas (2013), Bedi et al. (2014) e Sachdeva, Dhir e Kumar (2016), exemplificam e demonstram como utilizar o contexto para elaboração de diferentes SR;
- b) sistemas de recomendação: Bedi e Agarwal (2012), Gallego et al. (2013), Gallego, Woerndl e Huecas (2013), Bedi et al. (2014), Yao et al. (2014), Salman et al. (2015), Mashal, Alsaryrah e Chung (2016) e Sachdeva, Dhir e Kumar (2016), utilizam de diversas abordagens de SR. Estes sistemas apresentam formas diferentes para gerar recomendações e de acordo com estas formas eles são classificados em diferentes grupos;
- c) usuários: nos artigos de Bedi e Agarwal (2012), Gallego et al. (2013), Gallego, Woerndl e Huecas (2013), Bedi et al. (2014), Yao et al. (2014), Salman et al. (2015), Mashal, Alsaryrah e Chung (2016) e Sachdeva, Dhir e Kumar (2016), são mostrados SR que foram desenvolvidos para fornecer sugestões de diferentes funcionalidades para diferentes usuários, portanto, ao interpretá-los foi constatado que os usuários devem ser definidos entre conhecidos e não conhecidos; e
- d) aprendizado de máquina: Bedi e Agarwal (2012), Gallego, Woerndl e Huecas (2013), Bedi et al. (2014), Yao et al. (2014), Salman et al. (2015), Mashal, Alsaryrah e Chung (2016) e Sachdeva, Dhir e Kumar (2016) utilizam de diversos métodos de aprendizado de máquina para determinar quais sugestões os SR devem fornecer. Estas sugestões devem atender de forma satisfatória as necessidades de seus usuários, para isso diversas características devem ser levadas em consideração ao utilizar destes aprendizados.

Em resumo, os artigos supracitados auxiliaram a determinar os principais conceitos que caracterizam diferentes SPR. Através destes conceitos, uma série de características para

avaliação de Sistemas Proativos de Recomendação foram determinadas. Cada conceito é aprofundado nos capítulos posteriores e sintetizados através de ilustrações. Para cada ilustração é destacado os pontos mais relevantes que devem ser levados em consideração para a avaliação de um SPR.

3 CONTEXTO EM INTERNET DAS COISAS

Internet das Coisas (*Internet of Things - IoT*) alia tecnologias referente a diversas áreas de conhecimento. IoT é um conceito tecnológico que visa conectar objetos físicos a redes de comunicação, transformando-os em objetos virtuais que permitam o sensoriamento do mundo físico. Quando conectados, uma grande diversidade de dados e em grandes quantidades estarão disponíveis para desenvolvimento de diferentes sistemas que utilizam destes dados (ITU-T, 2012).

Contexto está relacionado com qualquer informação que possa ser utilizada para caracterizar uma situação e/ou necessidade de uma entidade, sendo que esta entidade pode assumir o papel de usuário, lugar e/ou objeto (DEY, ABOWD E SALBER, 2001).

Neste capítulo são mostrados quais aspectos influenciaram o nascimento do conceito IoT, quais as características que diferenciam a IoT da Internet convencional, as diferentes definições da IoT, os prognósticos baseados em projeções econômicas da IoT, definições para o termo Contexto e por fim como pode se utilizar a IoT para caracterizar diferentes Contextos.

3.1 ORIGENS E DEFINIÇÕES DA IOT

Embora a IoT venha ganhando mais notoriedade nos dias de hoje devido aos avanços tecnológicos em processamento de dados e capacidade de processamento, algumas ideias que corroboram para tal conceito, remetem ao final da década de 80 e começo da década de 90. No ano de 1991, Mark Weiser definiu o termo computação ubíqua como um conceito que prima tornar a interação entre homem e computador o mais natural possível, ou seja, os equipamentos tecnológicos recuam para o pano de fundo da vida rotineira dos seres-humanos, evitando animosidade entre os mesmos. Para tanto, considera-se duas etapas: inicialmente as interfaces devem ser desenvolvidas prevendo uma interação mais natural possível (como exemplo, as interações podem ser feitas através da utilização de gestos e fala), e por fim, deve ser previsto que a tecnologia usada seja sensível ao contexto em que ela está inserida (WEISER, 1991).

Antes de prosseguir com as definições de IoT, alguns termos que são comumente usados nesta área são definidos, sendo eles:

- a) dispositivos: são equipamentos que obrigatoriamente devem ter tecnologias de comunicação e opcionalmente podem ter elementos de sensoriamento, atuadores, determinação de dados, armazenamento de dados e processamento de dados (ITU-T, 2012);

- b) objetos: qualquer objeto presente no mundo físico (*physical things*) que foi feito para alguma finalidade em específico ou qualquer objeto presente no mundo da informação (*virtual things*) que serve para retratar alguma situação em específico (ITU-T, 2012); e
- c) coisas (*things*): qualquer objeto que tenha a capacidade de ser identificado e integrado a alguma rede de comunicação (ITU-T, 2012).

O termo IoT, foi cunhado pela primeira vez em uma palestra dada por Kevin Ashton em 1999. Em uma de suas publicações, Ashton justifica em qual acepção e contexto o termo IoT foi definido. Das principais razões apresentadas, destaca-se a seguinte:

O problema é que as pessoas têm tempo, atenção e precisão limitada – o que significa que não são tão boas em capturar dados e informações sobre certas coisas do mundo real... Se tivéssemos computadores que soubessem tudo o que há para saber sobre as coisas – através de dados que reunissem sem qualquer ajuda humana – seríamos capazes de encontrar e contar tudo e reduzir o desperdício, a perda e o custo de uma forma significativa. Gostaríamos de saber quando é que as coisas precisam ser substituídas, reparadas ou recordadas e se as coisas estão frescas ou se já passaram da validade (ASHTON, 2009).

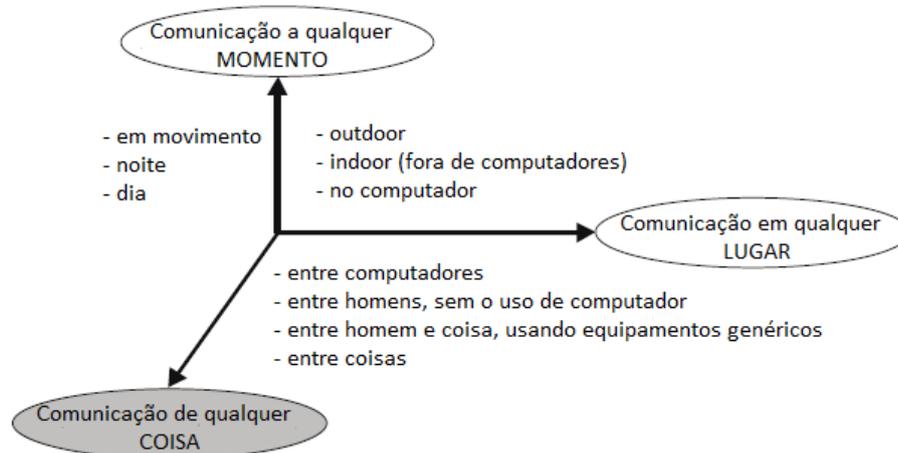
Fleish (2010), definiu alguns pilares fundamentais que auxiliam a estabelecer as diferenças entre a Internet usual e IoT. Tais características representam as expectativas que o conceito IoT deve contemplar, porém em alguns contextos não são mandatárias para a implantação da IoT. Tais comparações são: hardwares compactos contra espaçosos; trilhões contra bilhões de nós de rede; gargalo da *Last Mile* (termo utilizado para se referir ao meio físico que é usado para a comunicação entre provedores e usuários finais de infraestruturas de comunicações) contra taxas de dados maiores; não padronização contra padronização de identificação e endereçamento; uso centrado na máquina contra uso centrado no usuário e foco no sensoriamento ao invés da propaganda.

Segundo Perera et al. (2014), IoT permite que as coisas presentes no mundo se conectem com a Internet e comuniquem-se umas com as outras com o mínimo de intervenção dos seres humanos. Seu objetivo principal é criar um mundo melhor para a humanidade, onde as coisas saibam o que os seres humanos gostam, desejam, precisam e hajam de forma coerente sem instruções explícitas.

O órgão *International Telecommunication Union* – ITU apresenta uma definição para IoT no documento Y.2060 publicado no ano de 2012. Tal definição alude a uma infraestrutura global de informações provenientes de coisas. Tais coisas, englobam objetos presentes no mundo físico ou no mundo da informação que tenham a capacidade de serem unitariamente integradas as redes de comunicações existentes. As coisas originadas a partir dos objetos do mundo físico, devem ser capazes de serem detectadas (sensoriadas), incitadas (através de atuadores) e conectadas, como exemplos têm-se: equipamentos eletrônicos, robôs industriais, ambientes, entre outros. Para a outra vertente das coisas (originadas dos objetos do mundo das informações), estas devem ser capazes de serem armazenadas, processadas e acessadas, tais como conteúdos multimídias, aplicações de software, entre outras. Em ambas classes de coisas, critérios de segurança e privacidade devem ser assegurados (ITU-T, 2012).

Como mostrado na figura 3, a IoT incorpora o eixo “Comunicação de qualquer COISA” no gráfico de tecnologias de informações e comunicações, no qual já eram dispostos os eixos “Comunicação a qualquer MOMENTO” e “Comunicação em qualquer LUGAR” (ITU-T, 2012).

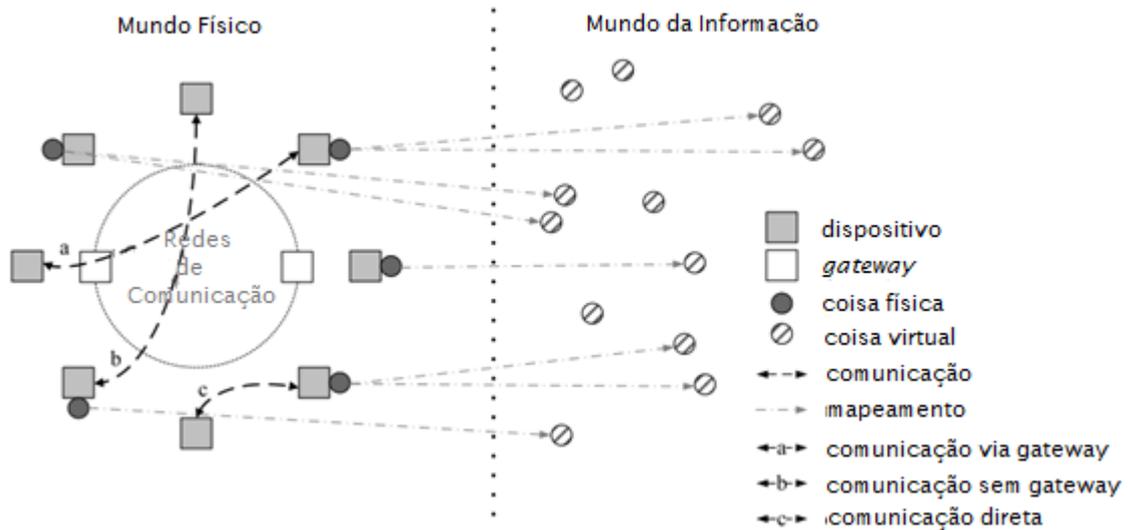
Figura 3 - Tecnologias de informações, comunicações e IoT



Fonte: Autor “adaptado de” ITU-T, 2012

Na figura 4, está representado uma visão técnica geral sobre o conceito da IoT.

Figura 4 - Visão técnica geral da IoT



Fonte: Autor "adaptado de" ITU-T, 2012

Em concordância com a recomendação Y.2060 elaborado pelo órgão ITU-T, um objeto no mundo físico pode ser modelado no mundo das informações como um ou mais objetos virtuais. Note que um objeto no mundo virtual pode existir mesmo não tendo como funcionalidade a representação de um objeto físico, tais como conteúdos multimídias, sendo que estes já foram feitos no mundo das informações (ITU-T, 2012).

Por ser uma tecnologia estudada por diversos pesquisadores, uma série de artigos apresentam definições que diferenciam entre si em certos aspectos. Algumas são apresentadas a seguir.

McEwen e Cassimally (2014) definiram IoT como um conceito envolvendo os seguintes elementos: objetos físicos, controle, sensores, atuadores e Internet. Um objeto desenvolvido para um propósito específico no mundo físico, quando este passa por um processo de modelagem virtual, ou seja, é inserido um dispositivo que permita-o captar informações, executar ações (através de sensores e atuadores respectivamente) e é integrado a uma rede de comunicações permitindo que este interaja com outros objetos para diferente finalidades, se faz a essência do conceito da IoT.

IoT surgiu a partir das diferentes tecnologias *wireless* desenvolvidas, onde tal conceito busca ampliar a quantidade de dispositivos que estão integrados com redes de comunicação, ou seja, ao invés de apenas computadores e *smartphones* estarem aptos a conectividade da Internet, diversos dispositivos devem ser desenvolvidos com a capacidade de se comunicarem e interagirem por meio da internet. Tais dispositivos devem ser identificados de forma única, devem ter a capacidade de compartilhar dados com outros dispositivos, além do que, devem

estar conectados a qualquer momento, em qualquer lugar, com qualquer coisa ou qualquer pessoa, de forma que possibilite a empregabilidade da Ubiquidade (BURANGE; MISALKAR, 2015).

No artigo de Asghar e Mohammadzadeh (2015) é dito que a IoT é tido como uma inovação disruptiva, onde foi emergida inicialmente de sensores, atuadores e dispositivos do dia a dia do ser humano que utilizavam de comunicações *wireless*. Tal conceito tem como característica fornecer métricas para mensurar o ambiente onde é aplicada. Através da internet, bilhões de sensores irão capturar uma enorme quantidade de dados que deverão ser analisados, interpretados e utilizados, de tal forma que o contexto poderá ser aferido dinamicamente através da modelagem e armazenamento dos dados gerados.

Em sua pesquisa, Riggins e Wamba (2015) mostra que do ponto de vista de grupos de pesquisas que utilizam de tecnologias RFID para desenvolvimento de projetos, IoT é definida como uma rede de coisas interconectados no mundo inteiro baseado nos protocolos padrões de comunicação, sendo que tais coisas devem ser endereçadas individualmente. Através da geração de enormes quantidades de dados, IoT proporcionará a base para o desenvolvimento da computação Ubíqua. Tais dados devem ser analisados e utilizados conforme o objetivo de quem faz esta análise.

Outra definição diz que a IoT é um conceito computacional que compreende objetos físicos do dia a dia conectados à internet. Estes objetos são capazes de se comunicar entre eles sem a interferência do ser humano, para tanto, tecnologias como *Radio-frequency identification (RFID)*, *Quick Response codes (QR)* e tecnologias *wireless* são utilizadas (SINGH, S.; SINGH, N, 2015).

Já na pesquisa de Al-Fuqaha et al. (2015), diz que a primeira fase da IoT teve seu início nas revoluções tecnológicas atuais referente aos conceitos de internet, dispositivos de comunicação móveis e *machine-to-machine (M2M)*. A IoT utiliza das últimas tecnologias desenvolvidas de *RFID*, *smart sensors*, *communication technologies* e protocolos da internet para integrar diferentes dados provenientes de diferentes coisas. A partir daí, tem-se que a premissa básica para caracterizar o conceito de IoT é: compreender sensores inteligentes, que colaboram diretamente para a origem de novas classes de aplicações sem o envolvimento do ser humano.

Após apresentada algumas definições de IoT, a que é utilizada para o desenvolvimento desta dissertação é dada por Asghar e Mohammadzadeh (2015) e de forma sucinta diz que a IoT fornece métricas para mensurar o ambiente de forma dinâmica onde está aplicada. Note que em algumas definições se faz necessário que as coisas estejam conectadas por meio da internet para que elas estejam inseridas no conceito de IoT e em outras definições, para que as coisas

estejam inseridas no conceito de IoT se faz necessário que elas apresentem apenas alguma forma de comunicação.

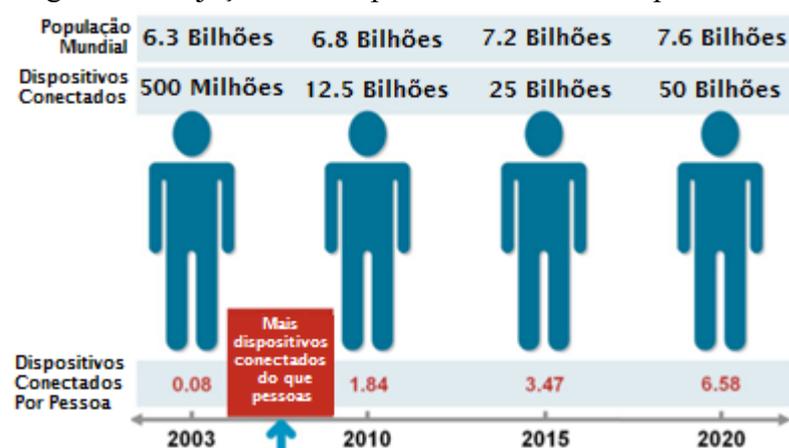
3.2 PERSPECTIVAS E GRUPOS DE APLICAÇÃO

A IoT cada vez mais tem estado presente no dia a dia das pessoas, portanto as perspectivas em relação a quantidade de dispositivos inseridos na IoT e os impactos econômicos que serão originados a partir deste novo conceito de tecnologia, estão crescendo e alavancando novos estudos e inovações na área. Devido a IoT estar sendo aplicada com diferentes finalidades, surgem diferentes grupos de aplicação que apresentam características próprias de como utilizar a IoT para atingir seus objetivos. Estes grupos são apresentados posteriormente nesta subseção.

Oferecendo grandes oportunidades de mercado, a IoT concentra-se principalmente em três grandes esferas, sendo elas manufatura de equipamentos, provedores de serviços para Internet e desenvolvedores de aplicativos (AL-FUQAHA et al., 2015).

De acordo com Evans (2011), no ano de 2020 mais de 50 bilhões de dispositivos que contemplam os conceitos da IoT estarão presentes no planeta, conforme demonstrado na figura 5. Dividindo tal número pela quantidade estimada de pessoas no planeta, chega-se ao número de 6,58 dispositivos conectados por pessoa.

Figura 5 - Projeções de Dispositivos Conectados por Pessoa



Fonte: Autor “adaptado de” Evans, 2011

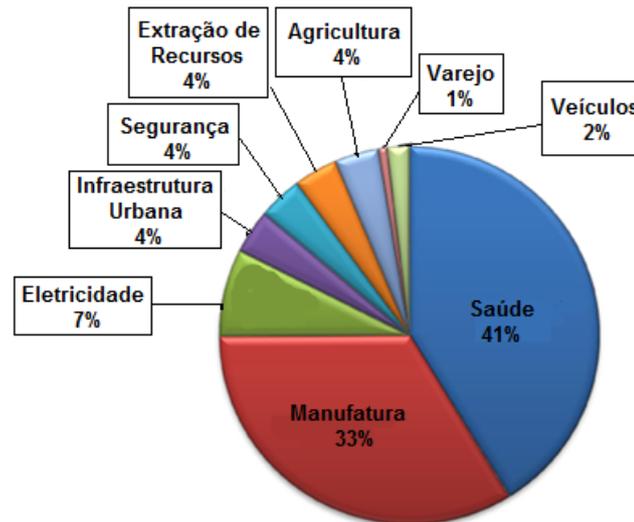
A empresa *Libelium*, separa em doze grupos distintos as aplicações que são desenvolvidas para a IoT, sendo elas: *smart cities*, *smart enviroment*, *smart water*, *smart metering*, *security and emergencies*, *retail*, *logistics*, *industrial control*, *smart agriculture*,

smart animal farming, domotic & home automation e ehealth. Para cada aplicação, uma série de dispositivos são utilizados. Estes grupos de aplicação já apresentam implementação prática, destacando-se:

- a) *smart cities*: a região *Castellón de la Plana* pertencente a Espanha, compreende 25 tipos de sensores que aferem e controlam dados de diferentes dispositivos. Estes dados são utilizados para apresentar diagnósticos da região e como indicativos para possíveis melhorias (IOTSSENS, 2016);
- b) *smart environment*: um projeto implementado na região norte da Espanha, faz o gerenciamento de alertas e avisos prévios sobre incêndios florestais através de diferentes dispositivos da IoT (SOLOBERA, 2010);
- c) *smart water e smart farming*: dispositivos estão sendo utilizados para fazer sensoriamento e controle de água em criadouros de peixes no Vietnã (LIBELIUM, 2016a);
- d) *smart metering*: dispositivos de medições de acelerações são colocados em carros de montanhas russas, afim de determinar forças de inércia exercidas nestes carros (BIELSA, 2012a);
- e) *security & emergencies*: dispositivos são responsáveis por medir níveis de radiação e através de tecnologias de comunicação enviam estes dados para estações de análise. Tais dados permitem detectar vazamentos de forma mais ágil (GASCÓN; YARZA, 2011);
- f) *retail e logistics*: diversos sensores são utilizados para auxiliar nas condições de armazenamento e entrega de produtos de padarias. Estas medições permitem entregas mais eficientes e melhor condicionamento dos produtos (LIBELIUM, 2015a);
- g) *industrial control*: dispositivos que tenham a capacidade de sensoriamento e comunicações de dados, são utilizados afim de permitir controle de carros de forma remota (BIELSA, 2012b);
- h) *smart agriculture*: através das medições provenientes de diversos dispositivos, sistemas de predição são desenvolvidos para determinar ações que devem ser tomadas afim de aumentar a qualidade de vinhas para fabricação de vinhos (LIBELIUM, 2015b); e
- i) *eHealth*: dispositivos são utilizados para captar dados provenientes de equinos. Estes dados são analisados afim de determinar possíveis comportamentos anormais e/ ou doenças (LIBELIUM, 2015c).

Ainda em relação as áreas de aplicações em IoT, na figura 6 são apresentados 9 grupos distintos de aplicação e a estimativa em porcentagem que apresentarão maiores implementações (AL-FUQAHA, 2015).

Figura 6 - Porcentagens das áreas em IoT
**Porcentagens do Potencial Econômico
 das Áreas de Aplicações da IoT**



Fonte: Autor “adaptado de” Al-Fuqaha, 2015

Em suma, por existir diferentes áreas de aplicação em IoT, grandes quantidades e diversidade de dados serão originados e auxiliarão para o desenvolvimento de diferentes sistemas com diferentes finalidades.

3.3 CONTEXTO

IoT através das coisas, permite que grandes quantidades de dados de diferentes tipos possam ser gerados. Estes dados quando analisados podem prover informações importantes sobre contextos onde as aplicações e sistemas estão. Porém, antes de prosseguir com as formas em que o contexto é obtido através da IoT, é necessário definir o conceito de Contexto.

De acordo com Dey, Abowd e Salber (2001), o termo Contexto é definido como: “Qualquer informação que possa ser usada para caracterizar a situação de uma entidade (i.e., pessoa, lugar ou objeto) que é considerada relevante na interação entre um usuário e uma aplicação, incluindo informações dos próprios usuários e aplicações...”

Do ponto de vista comportamental, Bazire e Brézillon (2005) apresentam algumas definições para contexto, destacando-se duas em especial. A primeira diz que “Contexto é

qualquer informação que caracterize uma situação relacionada a interação entre humanos, aplicações e o ambiente onde estão inseridos”. A segunda retrata que “Contexto é qualquer informação que possa ser usada para caracterizar e interpretar uma situação onde um usuário interage com uma aplicação em um determinado espaço de tempo”.

No dicionário Michaelis, a palavra Contexto apresenta os seguintes significados: “Conjunto de circunstâncias inter-relacionadas de cuja tessitura se depreende determinado fato ou situação; circunstância(s), conjuntura, situação”, “Conjunto de circunstâncias que envolvem um fato e são imprescindíveis para o entendimento deste”, “Encadeamento de ideias ou conjunto de circunstâncias que precedem ou se seguem a determinados elementos e pressupostos de um texto, aprofundando-se o significado quando de sua leitura ou análise” e “Conjunto de circunstâncias que envolvem uma palavra, uma frase ou um segmento de texto, e das quais podem depender seu significado e valor; encadeamento do discurso; composição, trecho, trama”. (MICHAELIS, 2015).

Como este trabalho aborda Sistemas Proativos de Recomendação e devido a definição de Dey, Abowd e Salber (2001) tratar de forma geral as informações que caracterizam um Contexto, esta é utilizada como definição padrão nesta dissertação.

No artigo de Perera et al. (2014), é apresentado que Contexto contempla uma série de eventos inter-relacionados, que quando iniciados determinam as condições deste Contexto. Eventos são separados entre duas categorias, sendo elas discretos e contínuos. Dado uma unidade de medida de intervalo de tempo p , estas duas categorias são diferenciadas entre si da seguinte forma:

- a) eventos discretos: é dado quando uma instância no momento t apresenta um estado de evento e no momento $t+p$ esta mesma instância apresenta um estado de evento distinto, como exemplo luz ligada em t e luz desligada em $t+p$;
- b) eventos contínuos: é dado quando uma instância no momento t apresenta um estado de evento e no momento $t+p$ continua no mesmo estado, como exemplo está chovendo em t e continua chovendo em $t+p$.

Em relação aos dados e/ou informações obtidos do Contexto, estes são classificados de acordo com a forma com que são obtidas, isto é: dados e/ou informações que são obtidos diretamente de sensores que captam eventos, sem que tenham sido originados a partir de dados e/ou informações de outros sensores, são classificados como caracterização de Contexto Primário (como exemplo localização obtida diretamente de um sensor GPS). Já dados e/ou informações que são obtidos a partir de inferências utilizando dos dados e/ou informações do

Contexto Primário, são classificados como caracterização de Contexto Secundário (como exemplo calcular a distância entre dois lugares distintos utilizando a posição de dois sensores GPS), estes dados e/ou informações geralmente são obtidos através de métodos computacionais. Na figura 7, estão dispostos alguns exemplos que caracterizam a diferença entre dado e/ou informações do contexto primário e secundário (PERERA et al., 2014).

Figura 7 - Contexto Primário e Secundário

		Categorias de Contexto (Perspectiva Operacional)	
		Primário	Secundário
Categorias de Contexto (Perspectiva Conceitual)	Local	Dados de localização vindo de sensores GPS (e.g. longitude e latitude)	Determinar distância através de dois GPS diferentes Imagem de um mapa obtido através de um provedor de mapas
	Ident.	Identificação de usuário através de RFID	Obtenção de lista de amigos através do Facebook Identificação facial através de sistemas de reconhecimento
	Tempo	Horário obtido de um relógio	Determinação da estação do ano baseado nos dados do clima Predição do horário baseado na atividade atual e calendário
	Atividade	Identificação de porta abrindo através do sensor de porta	Predição da atividade do usuário baseado em sua agenda Determinar a atividade do usuário através dos sensores de seu celular

Fonte: Autor “adaptado de” Perera et al., 2014

Para determinação de quais dados e/ou informações são utilizados para a representação do contexto, existem diversas abordagens que definem quais dados e/ou informações utilizar.

Schilit et al. (1994), apresenta três questões importantes que devem ser respondidas para caracterizar o contexto que se deseja analisar. A primeira pergunta se refere a localização onde a entidade que se deseja caracterizar o contexto está. Esta pergunta deve ser respondida com os dados e/ou informações sobre a localização da entidade. A segunda pergunta é relacionada a quem é a entidade. Esta pergunta deve ser respondida com dados e/ou informações sobre a entidade que está inserido no contexto em análise. Por fim, quais recursos estão próximos da entidade. Esta pergunta deve ser respondida com dados e/ou informações sobre os recursos disponíveis onde está localizada a entidade.

Abowd e Mynatt (2000), ao invés de caracterizar a representação do contexto em três perguntas eles utilizam cinco, sendo elas:

- a) quem interage: identifica qual ou quais elementos interagem de forma explícita ou implícita no ambiente de estudo. Note que tais elementos podem assumir o papel de um ou mais usuários de um SPR. Para demonstrar a importância de tal característica, tem-se que o comportamento do ser humano se torna maleável quando o mesmo está na presença de outras pessoas;
- b) com o que interage: diagnostica quais artefatos são usados para a interação entre os elementos e o ambiente. Transportando tal ponto para um ambiente de SPR, ao determinar quais itens são utilizados para interação entre os usuários e o sistema, ao longo do tempo e conforme o uso de tais itens, é possível mapear as atividades executadas dentro deste contexto;
- c) onde é feita a interação: relacionado as características do espaço físico em que se encontra o objeto de estudo. Tratando-se de SPR, ao identificar as características do ambiente em que este está submerso, é possível determinar a finalidade da recomendação a ser feita, ou seja, dado um usuário localizado em uma biblioteca a chance do mesmo necessitar de recomendações referente a livros é extremamente alta;
- d) quando é feita a interação: métricas referente ao tempo, são utilizadas para caracterização do contexto. Tais métricas são aplicadas aos elementos de tal contexto com diferentes objetivos. Do ponto de vista de SPR, ao contemplar a quantidade de tempo ou a estação do ano em que o usuário apresenta uma ou mais interações com algum item em específico, esta métrica pode servir como indicativo de preferência; e por fim
- e) porque é feita a interação: compreende as características que ajudam a determinação de quais razões levaram um elemento fazer uma interação em específico. Exemplificando, ao se aproximar da época de provas e devido as notas baixas em uma matéria em específico de um aluno qualquer, um SPR que compreende tal situação, faz sugestões de livros que possam auxiliar este aluno na melhoria de suas notas.

3.4 CONTEXTO EM IOT

Na IoT, sensores permitem a geração e obtenção de vários dados e de diferentes tipos. Através da interpretação individual e/ou conjunta dos dados originados e obtidos pelos dispositivos inseridos na IoT, sistemas desenvolvidos podem utilizar destes dados para diversos usos, tal como determinação do contexto onde estão inseridos. Nesta subsecção, será explicado o que são Redes de Sensores, definições do conceito Contexto e como a IoT fornece meios para obtenção de dados que ajudam a determinar um contexto em específico.

3.4.1 Redes de sensores em IoT

Redes de sensores em IoT são responsáveis pela geração, obtenção e comunicação de dados do mundo físico. As redes de sensores são componentes essenciais em IoT. Ao uni-las, além de surgir a capacidade de acesso remoto destes sensores, é possível utilizar de dados heterogêneos para desenvolver diferentes sistemas, como exemplo: monitoramento de ambientes, detecção de eventos, entre outros (ALCARAZ et al., 2010; PERERA et al., 2014).

Note que se tratando de SPR, os dados provenientes das redes de sensores inseridas na IoT, permitem que estes sistemas proativos possam determinar, através de métodos de aprendizado de máquina, momentos de pertinência para alguma sugestão em específico e quais são as preferências e/ou necessidades de seus usuários. Neste trabalho, as redes de sensores na IoT auxiliam na determinação do contexto onde está o usuário e/ou sistema e é tratada de forma indireta pelo modelo de avaliação de Sistemas Proativos de Recomendação Sensíveis ao Contexto quando analisado e avaliado a estrutura Internet das Coisas e Contexto.

De um ponto de vista técnico, estas redes são constituídas por um ou mais sensores, podendo comunicar-se entre si utilizando de diferentes protocolos e tecnologias *wireless* e/ou *wired*. Os sensores inseridos nestas redes podem ser classificados como homogêneos ou heterogêneos, sendo que uma rede típica de sensores é estruturada em 6 camadas: *Low-end Sensor Nodes*, *High-end Sensor Nodes*, *Sink Nodes*, *Low-end Computational Devices*, *High-end Computational Devices* e *Cloud* (PERERA et al., 2014).

Atualmente, grande parte dos sensores desenvolvidos para integração na IoT, utilizam de tecnologias *wireless*. Estas tecnologias são separadas em diferentes grupos de acordo com o alcance que têm, sendo elas: *wireless personal area network – WPAN*, *wireless local area network – WLAN*, *wireless metropolitan area network – WMAN*, *wireless wide area network – WWAN* e *satellite network* (ALCARAZ et al., 2010; PERERA et al., 2014).

Através das Redes de Sensores inseridas na IoT, dados de diversos tipos são coletados e utilizados em diferentes aplicações. Nesta etapa, é importante distinguir dois tipos de informações que são ligados diretamente aos dados obtidos (SANCHEZ et al., 2006; PERERA et al., 2014):

- a) dados: são dados brutos obtidos diretamente dos sensores e ainda não foram processados afim de um objetivo em específico;
- b) informações de contexto: são gerados através do processamento de dados brutos e através deste processamento determinam a coerência e/ou atinjam um objetivo em específico.

Destaca-se que diversas aplicações com diferentes finalidades estão inseridas na IoT, portanto, a IoT não se limita em apenas coletar dados de diferentes sensores e disponibiliza-los a estas aplicações, mas sim permitir que os dados coletados possam ser tratados e auxiliar na origem de outros tipos de inferências. Como exemplo, dada uma empresa que fabrica sensores de pressão que são destinados para determinação da qualidade da estrutura de pontes construídas para travessia, os dados provenientes destes sensores, quando inseridos na IoT, podem ser agrupados com dados de outros sensores e utilizados para determinar as condições de trânsito. Esta situação demonstra que a IoT compreende uma grande quantidade e diversidade de dados, permitindo que diferentes aplicações possam ser desenvolvidas utilizando de diferentes dados provenientes do Contexto onde são aplicadas (PERERA et al., 2014).

3.4.2 Sensibilidade ao contexto através da IoT e seus aspectos

Conforme dito anteriormente nesta dissertação, SPR estão inseridos em diferentes contextos e devem compreender diferentes âmbitos de aplicação, sendo que cada uma tem características próprias e níveis de urgências distintos. Sendo assim, ao se tratar de análises de contextos e dentro do paradigma da IoT, Perera et al. (2014) identificou três características que Sistemas e/ou Aplicações Sensíveis ao Contexto apresentam, sendo elas:

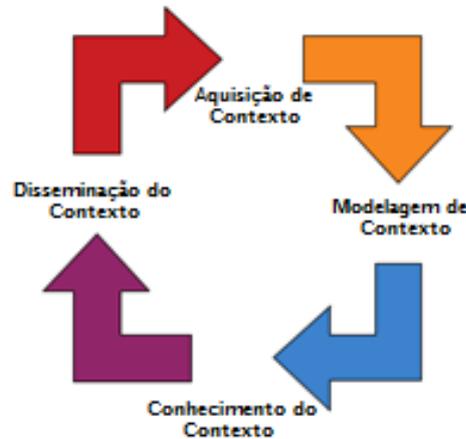
- a) *presentation*: o contexto pode ser usado para determinar quais informações e serviços devem ser apresentados ao usuário. Como exemplo, considere o cenário de um ambiente inteligente onde é possível identificar que um usuário entrou em um supermercado e está com seu *smartphone* em mãos. Dado uma aplicação *Mobile* Sensível ao Contexto que consiga se conectar com aparelhos de sua cozinha, tal como uma geladeira inteligente, uma lista de compras pode ser fornecida para este usuário

com itens que devem ser adquiridos. Esta situação exemplifica como o contexto é a base para prover informações que são úteis para o usuário. Este cenário é plausível quando estas aplicações são desenvolvidas dentro dos paradigmas da IoT, sendo que estes paradigmas são: prover qualquer serviço, a qualquer tempo, em qualquer lugar e para qualquer pessoa utilizando de qualquer rede de dados disponível;

- b) *execution*: execução de serviços de forma autônoma (proativa) caracteriza e é considerado um ponto crítico das aplicações desenvolvidas dentro dos paradigmas da IoT. Como exemplo, dado um cenário *Smart Home*, quando um usuário começa a dirigir de volta para sua residência, os serviços e aplicações da IoT inseridos nesta residência, iniciam de forma automática o acionamento do ar condicionado e da máquina de fazer café afim de já estarem prontos quando o usuário chegar. Este cenário evidencia que ações devem ser tomadas automaticamente de acordo com as informações do contexto e que comunicação *machine-to-machine* é uma parte significativa da IoT; e
- c) *tagging*: dentro do paradigma da IoT, existe uma extensa quantidade de sensores inseridos em inúmeros objetos presentes no dia a dia. Consequentemente, estes objetos produzem uma extensa quantidade de dados que necessitam ser coletados, analisados, fundidos e interpretados. Dados produzidos por apenas um sensor, não auxiliam para uma compreensão completa do contexto onde está inserido. Por outro lado, dados produzidos por diversos sensores e quando estes são interpretados conjuntamente, permitem uma compreensão mais completa do contexto onde estão inseridos. Este cenário demonstra que coleta, análise, fusão e interpretação de dados provenientes de diferentes sensores, executam um grande papel em Sistemas Sensíveis ao Contexto.

Dados e/ou informações do Contexto são gerenciadas através de Ciclos de Vida, isto é, etapas com diversas finalidades são executadas afim de se atingir um objetivo para os dados e/ou informações provenientes do Contexto. Perera et al. (2014), realizou diversas pesquisas de Ciclos de Vida de Contexto e identificou uma específica em seu trabalho que atende as necessidades para um bom gerenciamento de Contexto, levando em conta os dados e/ou informações obtidas através da IoT. Este ciclo é apresentado na figura 8.

Figura 8 - Ciclo de Vida das análises de Contexto



Fonte: Autor “adaptado de” Perera al., 2014

Resumidamente, têm-se que cada etapa é responsável pelas seguintes funcionalidades (PERERA et al., 2014):

- a) aquisição de contexto: é responsável por adquirir dados provenientes do contexto. De acordo com as formas que os dados são obtidos, estes são classificados em grupos diferentes:
- baseado na responsabilidade: o sensor físico ou virtual envia seus dados periodicamente ou em um instante específico para o software que é responsável pela coleta de dados do contexto (situação denominada de *push*), ou quando o software requisita dados do contexto através dos sensores de forma periódica ou em um instante específico (situação denominada de *pull*);
 - baseado na frequência: quando uma ação dentro do contexto ocorre e não se prolonga durante o tempo (como exemplo abrir a porta) é caracterizado como *instant events*. Quando uma ação dentro do contexto ocorre e se prolonga durante um tempo (como exemplo estar chovendo), é caracterizado como *interval events*;
 - baseado na fonte: é classificado de acordo com a fonte onde é obtida dados do contexto, sendo que podem ser adquiridas diretamente dos sensores (*sensor hardwares*), através de *middleware infrastructure* como exemplo redes globais de sensores (GSN – *Global Sensor Network*) e de outros locais de armazenamento de contexto (*context servers*), tais como base de dados, *web services*, entre outros;
 - baseado nos tipos de sensores: é classificado de acordo com o tipo de sensor para aquisição de dados do contexto. *Physical sensors* são sensores tangíveis

que geram dados por eles mesmos, como exemplo sensores de temperatura, luminosidade, entre outros. *Virtual sensors* retêm dados do contexto de outras fontes, geralmente *web services* (e.g. *twitter statuses*, *e-mail*, entre outros), note que *virtual sensors* não são tangíveis. *Logical sensors* utilizam da combinação dos dados provenientes do *physical sensors* e *virtual sensors* para obtenção de informações mais significativas do contexto;

- baseado em processo de aquisição: relacionado com o processo de aquisição de dados do contexto, sendo que é dividido em três formas. *Sense*, os dados são adquiridos diretamente dos sensores. *Derive*, informações são geradas através de operações simples (e.g. requisição de dados da *web*) ou através de operações mais complexas (e.g. funções matemáticas). *Manually provided*, os usuários fornecem dados manualmente do seu contexto através de configurações pré-definidas (e.g. um usuário diz que não gosta de receber notificações no período da tarde).
- b) modelagem de contexto: é responsável pela determinação de quais dados e/ou informações são utilizados para a representação do contexto. Como exemplo, Schilit et al. (1994), utiliza três questões que auxiliam na escolha de quais dados e/ou informações utilizar para fazer a representação do contexto, sendo elas: onde você está, quem é você e quais recursos estão próximos. Em Perera et al. (2014), seis técnicas são identificadas e utilizadas para a manipulação dos dados e/ou informações que fazem a representação do contexto: *key-value*, *markup schemes*, *graphical*, *object based*, *logic based* e *ontology based modelling*. Tipicamente existem duas tratativas que devem ser feitas em qualquer modelo de representação de contexto:
- processo de modelagem do contexto: dados e/ou informações novas do contexto necessitam ser definidas em termos de atributos, características, relação com os dados e/ou informações necessárias para determinação do contexto, qualidade dos atributos deste contexto e *queries* que fazem as sincronizações para requisição dos sensores;
 - organização do contexto de acordo com o modelo: o resultado da representação do contexto é validado nesta tratativa, ou seja, os novos dados e/ou informações são combinados e adicionados as informações que já estão armazenadas para representação do contexto. Depois da validação, os dados e/ou informações ficam disponíveis quando requisitados para uso.

- c) conhecimento do contexto: através dos dados e/ou informações validadas na etapa anterior, métodos de aprendizado são utilizados nesta fase para obtenção de novos conhecimentos a respeito do contexto. As técnicas mais utilizadas para *context reasoning decision models*, são as de Inteligência Artificial e Aprendizado de Máquina, destacando-se: *Decision Tree, Naive Bayes, Hidden Markov Models, Support Vector Machine, k-Nearest Neighbor e Rules*. Esta etapa é aplicada afim de sanar duas deficiências de dados obtidos do contexto: imperfeição, ou seja, quando se tem erros, imprecisões, ambiguidade e ausência de dados e/ou informações que estão dispostos para caracterizar o contexto e incertezas, que se refere a confiabilidade dos dados e/ou informações obtidos. Esta etapa pode ser dividida em três fases:
- pré-processamento do contexto: esta fase faz a filtragem e manipulação dos dados provenientes dos sensores. Caso os sensores apresentem problemas e resultem em valores de dados imprecisos ou ausentes, neste caso, algumas tratativas podem ser executadas, tais como: preenchimento dos valores desconhecidos, tratamentos dos *outliers*, validação dos dados através de comparação com outras fontes, entre outras;
 - fusão de dados dos sensores: combina dados de diferentes sensores afim de produzir informações do contexto mais apuradas, confiáveis e gerar informações que não poderiam ser obtidas apenas de um sensor. Na IoT, esta é uma fase muito importante, devido a existência de bilhões de sensores que irão gerar diferentes tipos de dados; e
 - inferências do contexto: Gera informações de alto-nível do contexto a partir de dados de baixo-nível do contexto (dados provenientes dos sensores).
- d) disseminação do contexto: compreende os métodos de envio das informações de contexto para os consumidores destas informações. Da perspectiva do consumidor de informações de contexto, esta etapa apresenta as mesmas características que a etapa Aquisição do Contexto e mais dois métodos:
- *query*: o consumidor faz requisições de informações do contexto no formato de *query* para o sistema gerenciador de contexto, que por sua vez usa esta *query* para produzir resultados; e
 - *subscription*: o consumidor de informações do contexto estabelece condições para o sistema gerenciador de contexto para envio de informações, podendo ser de forma periódica e/ou quando alguma condição é violada.

Em suma, ao avaliar e/ou desenvolver um Sistema Proativo de Recomendação, algumas questões importantes devem ser conhecidas. De acordo com o explicado nesta seção, dados e/ou informações que devem ser utilizados para definir um Contexto em específico, são aqueles que auxiliam na caracterização de uma entidade (neste caso o sistema e/ou usuário). As perguntas pré-definidas que auxiliam na caracterização do contexto, são: quem interage, com o que interage, onde interage, quando interage e porque interage. Dentro do paradigma da IoT, estas perguntas podem ser respondidas através de dispositivos inseridos em diferentes redes de sensores, de forma que dados e/ou informações primárias ou secundárias do contexto possam ser coletadas pelos SPR. Como informações do contexto se alteram ao longo do tempo, ciclos de vida para a tratativa destas informações devem ser empregadas para que SPR possam, de forma contínua, atender as necessidades de seus usuários. Portanto, quando se trata de SPR, dados e/ou informações do contexto devem ser coletados, tratados e gerenciados (através de ciclos de vida). Estes dados e/ou informações do contexto, auxiliam os SPR determinar quais itens, ações e/ou serviços recomendar e em quais momentos estas recomendações devem ser feitas de forma proativa.

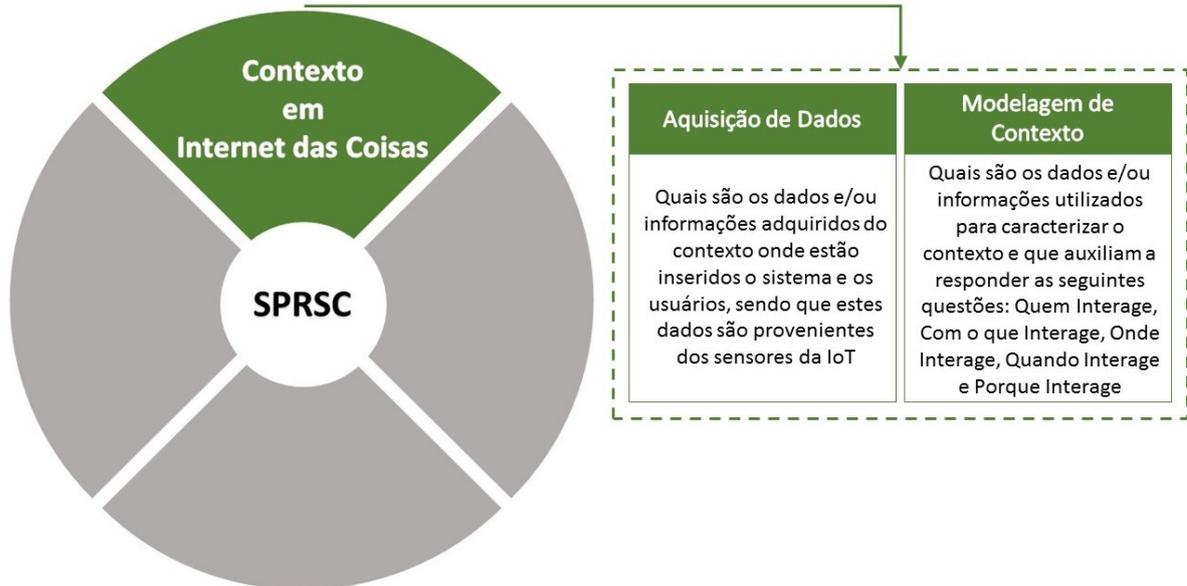
3.5 ESTRUTURA CONTEXTO EM IOT DE SPRSC

Com base nas pesquisas analisadas e conceitos explicados ao longo desta seção, a figura 9 foi desenvolvida nesta dissertação afim de auxiliar na avaliação de Sistemas Proativos de Recomendação quando os aspectos a serem avaliados estão relacionados a caracterização do contexto onde sistema e usuário estão inseridos. Conforme dito anteriormente, a IoT compreende redes de sensores que fornecem grandes quantidades de dados provenientes de diversos sensores inseridos em inúmeros objetos do dia a dia das pessoas. Note que estes dados são de diferentes tipos devido serem provenientes de diferentes objetos. Sendo assim, ao se avaliar um SPRSC em relação à forma com que é feita as tratativas do contexto, dois aspectos devem ser determinados. O primeiro é determinar quais são os dados que um SPRSC em específico pode adquirir do contexto através dos dispositivos inseridos na IoT. De acordo com a finalidade do sistema proativo, todos os dados disponíveis podem ser usados ou não. Portanto, o segundo aspecto está em determinar quais dados que são usados para a modelagem do contexto onde estão inseridos usuários e sistema e de que forma é feita a modelagem.

Vale ressaltar que a figura 9 desenvolvida nesta dissertação e que representa a estrutura Contexto em IoT do modelo de avaliação de SPRSC é embasada na análise de diferentes pesquisas relacionadas a Sistemas Proativos de Recomendação e IoT, destacando-se

principalmente as pesquisas de Bedi e Agarwal (2012), Gallego, Woerndl e Huecas (2013), Perera et al. (2014) e Salman et al. (2015). Através destas análises, fica claro que dados do contexto coletados através dos dispositivos inseridos na IoT, é um aspecto de suma importância para avaliação de sistemas proativos, devido estes dados serem a base para determinar quais momentos são propícios para uma sugestão autônoma e quais itens, ações e/ou serviços sugerir.

Figura 9 - Estrutura Contexto em IoT do modelo de avaliação de SPRSC



Fonte: Autor

4 SISTEMAS DE RECOMENDAÇÃO

As técnicas de Sistemas de Recomendação (SR) tradicionais são partes integrantes dos Sistemas Proativos de Recomendação Sensíveis ao Contexto (SPRSC). Para a avaliação dos SPRSC, é necessário que as abordagens mais tradicionais de SR sejam explicadas.

Este capítulo compreende os métodos mais comuns de SR, sendo eles classificados em cinco esferas distintas: baseado em conteúdo (*content-based methods*), filtragem colaborativa (*collaborative methods*), demográfico (*demographic*), funções de utilidade (*utility-based*) e conhecimento (*knowledge-based*). Os sistemas que compreendem duas ou mais das técnicas anteriores aplicados na elaboração de suas sugestões são denominados de sistemas híbridos (*hybrid methods*). Para cada esfera em individual é apresentado também suas vantagens e desvantagens.

4.1 SR BASEADOS EM CONTEÚDO

As sugestões feitas pelos SR Baseados em Conteúdo, são elaboradas a partir das descrições dos itens, cujo quais possuem características semelhantes a uma lista de itens de preferência de um usuário qualquer.

Tais sistemas, são extremamente utilizados em aplicações Web desenvolvidas para a interação de usuários. Normalmente, tais aplicações apresentam uma lista de itens e os usuários escolhem aqueles cujo qual possuem maior interesse. Como exemplos deste ambiente têm-se: sites especializados na apresentação de notícias, reportagens e na venda de livros. Em relação aos sites de notícias, em sua página inicial são apresentadas diversas manchetes, sendo que o usuário ao se interessar por uma em específico e clicar é habilitado a leitura completa da matéria. Note que, ao existir uma extensa gama de manchetes há a necessidade de se filtrar quais são mais apropriadas para os diversos usuários, ou seja, algumas serão mostradas e outras não. O SR Baseados em Conteúdo é responsável por escolher quais serão habilitadas para uma visualização prática. Para isto, o sistema se baseia nas características das notícias que são de preferência de um usuário em particular e aquelas que apresentarem aspectos semelhantes serão fornecidas. Em relação aos sites desenvolvidos para a venda de livros, um usuário em específico comprou um livro e gostou do mesmo, este livro apresenta uma sequência, então o SR sugere a sequência deste livro devido estes apresentarem aspectos semelhantes (PAZZANI; BILLSUS, 2007; LOPS; DE GEMMIS; SEMERARO, 2011).

Em concordância com Pazzani e Bilssus (2007), a ideia principal da técnica de sistema supracitada, é que tais sistemas são baseados em regras gerais, ou seja, recomendam outros produtos que compartilham de algumas características semelhantes aos já adquiridos e/ou de preferência dos usuários.

Do ponto de vista técnico, esta abordagem se fundamenta em determinar como são classificados os itens que estão disponíveis e contemplados pelo SR Baseados em Conteúdo e qual forma que será elaborado o histórico dos usuários deste sistema. Após formulada tais etapas, através de métodos de aprendizado e classificação, são elaborados os modelos de preferências dos usuários.

Em suma, estes métodos de classificação e aprendizado determinam uma função que caracteriza o perfil de interesses dos usuários. Uma vez determinada tal função, o SR Baseados em Conteúdo pode prever se um item novo deve ser recomendado ou não para um usuário em específico baseado em seu perfil. Para tal sistema apresentar sugestões satisfatórias é de *conditio sine qua non* que o mesmo contemple informações suficientes que permitam a distinção entre itens que um usuário em específico gosta e os que não gosta. Tal característica nem sempre é possível e/ou viável, sendo assim, muitas sugestões provenientes destes sistemas podem não atender as pretensões de seus usuários.

Na estrutura Sistemas de Recomendação do modelo estrutural SPRSC desenvolvido nesta dissertação, é utilizado esta técnica onde o perfil de preferências do usuário é determinado através de um algoritmo de aprendizado de máquina denominado C4.5.

As vantagens e desvantagens desta técnica de SR é demonstrada no quadro 1 e foram obtidas a partir das pesquisas de Pazzani (1999), Burke (2002) e Schafer et al. (2007).

Quadro 1 - Vantagens e Desvantagens SR Baseado em Conteúdo

(continua)

SR Baseados em Conteúdo	
Vantagens	Desvantagens
<p>Conhecimento do contexto em que o sistema está inserido não é necessário para gerar sugestões. Devido tal abordagem fundamentar a elaboração de suas recomendações nas avaliações provenientes de um usuário e nas características que descrevem um item, este não necessita de métodos que permitam o aprendizado do contexto em que se encontra.</p>	<p><i>Ramp-up</i> de novo usuário. As recomendações para que sejam prestativas e tenham qualidade, devem ser elaboradas a partir das diversas avaliações referentes aos itens de um usuário em específico. Ao ingressar um novo usuário no sistema, muitas vezes este não fez nenhuma avaliação sobre os itens, portanto o sistema não apresenta dados suficientes, resultando na elaboração de recomendações com pouca qualidade.</p>
<p>Conforme as interações são feitas pelos usuários o mesmo proporciona uma melhor adaptabilidade e consequentemente suas recomendações se tornam mais precisas e prestativas.</p>	<p>Tal abordagem depende de uma extensa base de dados das avaliações dos itens para que suas recomendações atendam com qualidade as expectativas dos usuários.</p>
<p>Métricas de coleta de avaliações implícitas podem ser aplicadas ao usuário, tais como: Quantidade de visualizações, análise de tempo, quantidade comprada, entre outras que não necessitam que o usuário avalie explicitamente um item em específico.</p>	<p>Sistemas Baseado em Conteúdo dependem das características explícitas associadas aos itens que podem ser recomendados, isto é, imagine que tal sistema faz recomendações de filmes, o mesmo precisa ter informações a respeito dos atributos que os qualifica, como atores, gênero, ano de lançamento, entre outros. Concluindo, tais sistemas ficam à mercê de bons detalhamentos dos itens.</p>
	<p>Como a sugestão é baseada nas características explícitas dos itens, aqueles que apresentam propriedades diferentes são desconsiderados, mesmo podendo ser do agrado do usuário, isto é, imagine que um usuário através de suas avaliações demonstrou forte interesse por vestuário esportivo, as recomendações provenientes serão limitadas a abordar apenas tais vestimentas, mesmo que o usuário se interesse também por vestimenta formal.</p>
	<p><i>Portfolio Effect</i>. Um sistema de recomendação ideal não deve sugerir itens repetidos ou que pertença ao usuário de alguma forma. Dependendo do contexto em que é aplicado o sistema, alguns itens podem remeter a mesma finalidade, porém com alguma característica ligeiramente diferente, fazendo o sistema se equivocar e recomendar um item que tenha a mesma finalidade. Como exemplo, tem-se: Sistemas que sugerem notícias de diferentes jornais. Muitas vezes, apesar de serem escritos ou intitulados de formas dispares, remetem a mesma ideia e opinião.</p>

Quadro 2 - Vantagens e Desvantagens SR Baseado em Conteúdo

(conclusão)

SR Baseados em Conteúdo	
Vantagens	Desvantagens
	Conforme um usuário em específico interage com o sistema através de suas avaliações, este desenvolve um perfil para o mesmo e conseqüentemente elabora suas recomendações. Caso o usuário tenha uma mudança brusca em suas preferências o sistema exige um tempo relativamente extenso para verificar que o perfil foi alterado, gerando recomendações errôneas e não satisfatórias. Um exemplo prático para tal circunstância é: Uma pessoa demonstra através de suas avaliações que é um apreciador de bebidas alcoólicas, por algum motivo o mesmo passa a não ingerir mais estas bebidas, portanto até que o sistema aprenda a respeito deste novo gosto irá fazer algumas sugestões inapropriadas

Fonte: Autor

4.2 SR COM FILTRAGEM COLABORATIVA

Baseado em hábitos sustentados durante décadas pelos seres humanos, tal como reuniões para discussão de opiniões sobre livros, a ideia principal para tais sistemas é: através de métodos de aprendizado de máquina, elaborar e validar sugestões a partir da opinião de diversos usuários. Exemplificando tal fundamento, imagine um grupo de pessoas que se reúnem ao longo de vários anos e discutem sobre filmes para assistirem. Com o passar dos anos e inúmeras reuniões, duas competências principais de aprendizado são desenvolvidas entre os integrantes do grupo. A primeira forma e mais simples se caracteriza em quanto mais integrantes falarem bem de um filme em específico maior será a probabilidade da sugestão ser boa e ao contrário, quanto mais integrantes avaliarem de forma negativa, a probabilidade maior é de que o filme não agrade. A segunda competência e mais complexa para se desenvolver, está em distinguir entre os integrantes, quais são aqueles que possuem preferências semelhantes ou díspares, ou seja, sugestões provenientes dos integrantes terão peso de influência variável de acordo com o integrante á quem é feita a sugestão (BALABANOVIĆ; SHOHAM, 1997; SCHAFFER et al., 2007).

Ao integrar pareceres e avaliações de diversos usuários com a internet (como exemplo computadores), as opiniões de inúmeros usuários são consideradas, sendo que estas podem ser processadas em tempo real permitindo que não apenas aqueles itens que são os mais bem avaliados pela grande maioria sejam recomendados, mas também sugerir itens mais apropriados

para um grupo específico de usuários de acordo com seus interesses e características, gerando assim sistemas ainda mais personalizados (SCHAFER et al., 2007).

Em outras palavras, ao invés de basear as recomendações em itens que apresentam características semelhantes (SR Baseado em Conteúdo), estes sistemas baseiam a elaboração de recomendação em grupos de usuários semelhantes, ou seja, os itens que foram bem avaliados por um grupo em específico é recomendado aos usuários que se assemelham aos integrantes deste grupo (BALABANOVIĆ; SHOHAM, 1997).

Em concordância com SCHAFER et al. (2007), ao se desenvolver um SR com Filtragem Colaborativa, alguns aspectos do ambiente determinam se o mesmo será prestativo e eficiente. Tais propriedades são divididas em três domínios distintos. No quadro 2 estão explicados quais são os domínios e quais características intrínsecas de cada domínio.

Quadro 3 – Aspectos para desenvolvimento de SR com Filtragem Colaborativa

(continua)

Domínio	Características do Domínio
<p>Distribuição de Dados: as propriedades embutidas nesta categoria se referem a quantidade e a forma de como são tratados os dados nos sistemas.</p>	<p>Quantidade de itens: um sistema que contempla poucos itens é desnecessário, devido que o usuário pode aprender sobre todos sem a necessidade de ser auxiliado por um Sistema de Recomendação.</p> <p>Quantidade de avaliações por item: itens com poucas avaliações podem não apresentar informações suficientes para predições e sugestões satisfatórias.</p> <p>Quantidade maior de usuários do que itens a serem recomendados: se houver poucas avaliações provenientes por um único usuário, é necessário que o sistema compreenda inúmeros usuários.</p> <p>Usuários devem avaliar mais de um item: usuários que avaliam apenas um único item contribuem apenas para avaliações singulares, ou seja, não ajudam a caracterizar qual é o seu perfil de preferência, conseqüentemente não agrega valor para sistemas colaborativos.</p>
<p>Subjetividade de Classificação: as propriedades embutidas nesta categoria se referem a como as características dos usuários são tratadas pelos SR com Filtragem Colaborativa.</p>	<p>Para cada usuário da comunidade existem outros que compartilham das mesmas necessidades e/ou preferências: caso um usuário em específico tenha necessidades únicas e/ou características singulares, o mesmo não agregará valor nenhum para o sistema. Tais sistemas tem melhor aprendizado e conseqüentemente apresentam melhores sugestões quando os usuários do mesmo compartilham de preferências e necessidades semelhantes.</p> <p>Para sugestões e validações eficientes de itens, alguns critérios para preferências devem ser contemplados: SR com Filtragem</p>

Quadro 4 – Aspectos para desenvolvimento de SR com Filtragem Colaborativa

(conclusão)

Domínio	Características do Domínio
<p>Subjetividade de Classificação: as propriedades embutidas nesta categoria se referem a como as características dos usuários são tratadas pelos SR com Filtragem Colaborativa.</p>	<p>Colaborativa permitem que seus usuários compartilhem seus gostos uns com os outros. Para itens que são extremamente subjetivos, tal como músicas, ou quando tiverem diferentes critérios de características, tal como carros, oferecer opções que permitam que os usuários definam quais características os mesmos querem, podem auxiliar na validação do item pelo sistema. Lembrando que estes critérios devem ser utilizados com cuidado, pois se forem dispostos muitos, esta abordagem colaborativa se torna um sistema híbrido.</p> <p>Itens devem ser similares: os itens que estão sendo monitorado pelo sistema devem ser de grupo similares, sendo que apenas algumas características diferenciam entre si, ou seja, um usuário que anseia por sugestões de ferramentas espera recomendações de itens que constituem o grupo de ferramentas e não de artigos eletrônicos. Caso os usuários desejem receber sugestões diversificadas, o sistema poderá incluir itens diversificados em suas recomendações</p>
<p>Durabilidade dos Dados: as propriedades embutidas nesta categoria se referem a longevidade das informações contempladas pelos sistemas.</p>	<p>Durabilidade de um Item: devido às diferentes categorias de itens, é necessário determinar quais grupos se tornam obsoletos em um curto ou longo prazo de tempo ou aqueles que não se tornam obsoletos. Como exemplo, sistemas que fazem sugestões de notícias diárias, com o passar dos dias, tais notícias na maioria das vezes se tornam desatualizadas, portanto as mesmas devem ser retiradas do sistema.</p> <p>Durabilidade de Preferência: abordagens colaborativas têm maior êxito em ambientes que as preferências não se alteram rapidamente, porém é necessário que o sistema preveja quando tais preferências se modifiquem para que suas sugestões não se tornem ineficientes e dispensáveis</p>

Fonte: Autor

Os domínios supracitados, representam diretrizes para o desenvolvimento de SR com Filtragem Colaborativa, porém muitas vezes o objetivo final de um desenvolvedor não compreende todas, como exemplo, sistemas que têm como finalidade fazer sugestões de diversos grupos de objetos, não compreendem a característica de homogeneidade.

Em suma, SR com Filtragem Colaborativa abrangem inúmeros usuários. Aplicando métodos de aprendizado de máquinas, é determinado quais características são relevantes para a divisão destes usuários em grupos distintos. Uma vez classificados, esses serão base para elaboração de sugestões. Formalmente pode-se classificar tais sistemas da seguinte forma: as

recomendações feitas a um usuário específico são determinadas a partir de preferências de outros usuários que apresentam preferências semelhantes, ou seja, um grupo de usuários têm preferências de ações A e B, caso o usuário em questão tenha preferência por A o sistema assume que este pode também ter preferência por B.

As vantagens e desvantagens desta técnica de SR é demonstrada no quadro 3 e foram obtidas a partir das pesquisas de Pazzani (1999), Burke (2002) e Schafer et al. (2007).

Quadro 5 - Vantagens e Desvantagens SR Filtragem Colaborativa

SR com Filtragem Colaborativa	
Vantagens	Desvantagens
Cross-gênero ou “ <i>outside the box</i> ”. Abordagens colaborativas permitem que o sistema indique para um usuário em específico, produtos com características completamente diferentes, sendo que estas indicações são baseadas nas avaliações de outros usuários que possuem gostos semelhantes. Conforme dito neste capítulo, quando um usuário através de suas preferências é inserido em um grupo característico, o sistema passa a indicar produtos que foram bem avaliados por este grupo, mas ainda não foram utilizados por este usuário, fazendo com que a carta de itens a serem oferecidos aumentem.	<i>Ramp-up</i> de novo usuário. Explicado na abordagem Baseada em Conteúdo.
Conhecimento do contexto em que o sistema está inserido não é necessário para gerar sugestões. Devido tal abordagem fundamentar a elaboração de suas recomendações nas avaliações provenientes dos usuários e na correlação com outros usuários, esta abordagem não necessita de métodos que permitam o aprendizado do contexto em que se encontra.	<i>Ramp-up</i> de novo item ou <i>Early Rater Problem</i> . À medida que novos itens são inseridos no sistema, estes não apresentam avaliações. Como os itens a serem recomendados necessitam de inúmeras avaliações de outros usuários, este novo item precisará de outras técnicas para fazerem parte da carta de sugestões. Como exemplo: Sistemas que sugerem notícias diárias, ou seja, a cada dia novos informes são gerados e estes ainda não apresentam diversas avaliações.
Conforme as interações são feitas pelos usuários o mesmo proporciona uma melhor adaptabilidade e conseqüentemente suas recomendações se tornam mais precisas e prestativas.	<i>Gray Sheep Problem</i> . Abordagens colaborativas tem como intuito adequar seus usuários em grupos bem característicos conforme seus gostos. Para tal classificação, as avaliações provenientes dos usuários referente a diversos itens são utilizadas. Para aqueles usuários que possuem preferências bem singulares, ou seja, não compatíveis aos demais, estes não se adequam a nenhum grupo, portanto as recomendações feitas pelo sistema se tornam duvidosas e muitas vezes ineficientes.
Assim como a abordagem Baseado em Conteúdo, métricas de coleta de avaliações implícitas podem ser aplicadas ao usuário.	SR com Filtragem Colaborativa dependem de uma extensa base de dados das avaliações dos itens para que suas recomendações atendam com qualidade as expectativas dos usuários.

Fonte: Autor

4.3 SR DEMOGRÁFICOS

Baseado na abordagem Filtragem Colaborativa, tal SR classifica seus usuários em grupos distintos, porém ao invés de utilizar as avaliações e preferências para tal processo, é utilizado características de cunho pessoal.

Nesta abordagem, primeiro o sistema deve adquirir características de cunho pessoal de seus usuários, tais como: Idade, País, Estado, Cidade, Estado Civil, Raça, entre outros que auxiliam a descreve-los. Após feita esta fase de compreensão, os usuários são alocados em grupos distintos de acordo com suas características e as sugestões provenientes do sistema serão direcionadas de forma personalizada para tais grupos. Nota-se que os dados inseridos pelos usuários podem ser de formas estruturadas ou até mesmo não estruturadas. O algoritmo utilizado pelo sistema para a classificação de seus usuários, deve ser escolhido de acordo com a forma com a qual os dados dos mesmos são auferidos. Para a elaboração de sugestões, os algoritmos de uso comum entre os sistemas de recomendação podem ser utilizados (PAZZANI, 1999; BURKE, 2002; LIKA; KOLOMVATSOS; HADJIEFTHYMIADES, 2014).

Vale ressaltar que devido o sistema sustentar seu aprendizado e conseqüentemente sugestões em grupos que apresentem características de cunho pessoal semelhantes, questões de privacidade e veracidade podem tornar tal abordagem dificultosa e/ou inverossímeis.

As vantagens e desvantagens desta técnica de SR é demonstrada no quadro 4 e foram obtidas a partir das pesquisas de Pazzani (1999), Burke (2002) e Schafer et al. (2007).

Quadro 6 - Vantagens e Desvantagens SR Demográficos

(continua)

SR Demográficos	
Vantagens	Desvantagens
<p>Cross-genre ou “<i>outside the box</i>”. Semelhante à técnica Filtragem Colaborativa, abordagens demográficas permitem que o sistema indique para um usuário em específico, produtos com características completamente diferentes baseada nas avaliações de outros usuários pertencente ao mesmo grupo. Conforme dito anteriormente, quando um usuário através de suas qualidades de cunho pessoal é inserido em um grupo característico, o sistema passa a indicar produtos que foram bem avaliados por este grupo, mas ainda não fora utilizada por este usuário, fazendo com que a carta de itens a serem oferecidos aumente.</p>	<p><i>Gray Sheep Problem</i>. Assim como as abordagens colaborativas, tais sistemas têm como intuito adequar seus usuários em grupos bem característicos, porém neste caso, os elementos que são usados para tais categorizações são as características de cunho pessoal. Para aqueles usuários que possuem qualidades pessoais bem singulares, ou seja, não compatíveis aos demais, estes não se adéquam a nenhum grupo, portanto as recomendações feitas pelo sistema se tornam duvidosas e muitas vezes ineficientes.</p>

Quadro 7 - Vantagens e Desvantagens SR Demográficos

(conclusão)

SR Demográficos	
Vantagens	Desvantagens
Como o sistema se baseia nas características de cunho pessoal e essas podem ser auferidas de uma forma mais rápida, não existe a necessidade de analisar as preferências de um usuário em particular, portanto não existe o <i>Ramp-up</i> de novo usuário.	<i>Ramp-up</i> de novo item ou <i>Early Rater Problem</i> . Explicado na abordagem Filtragem Colaborativa.
Conforme as interações são feitas pelos usuários o mesmo proporciona uma melhor adaptabilidade e conseqüentemente suas recomendações se tornam mais precisas e prestativas.	A qualidade das recomendações provenientes do sistema depende de uma extensa base de recomendações provenientes dos usuários.
	Ao se desenvolver tal abordagem, técnicas que permitam a coleta de informações de características de cunho pessoal devem ser apreciados no desenvolvimento do sistema. Muitas vezes, tal dado pode não ser fornecido devido questões de privacidade do usuário, portanto requisitos de segurança de informações devem ser esclarecidos e implementados paralelamente.

Fonte: Autor

4.4 SR FUNÇÕES DE UTILIDADE

Em SR Funções de Utilidade, todos os itens são representados por funções de utilidade. Cada função, além de retratar características físicas e de funcionalidade do objeto, representa aspectos diversos provenientes das condições de estado, tais como: termos de garantia, item novo ou usado, valores de frete, parcelamento de compra, entre outros. Neste SR, o usuário, determina por meio de interfaces que o sistema proporciona, quais funções são mais importantes para ele no atual momento da interação. As sugestões são determinadas de acordo com o valor de utilidade atribuído para cada item pelo usuário, ou seja, cada item terá um grau de potencial variável para recomendação de acordo com a necessidade momentânea de um usuário em particular. Diferentemente dos SR Demográficos, SR Baseado em Conteúdo e SR Filtragem Colaborativa, SR Funções de Utilidade não dependem do armazenamento de inúmeras avaliações provenientes dos usuários para elaborar suas recomendações, pois através de uma interação inicial com o sistema é determinada quais funções de utilidade atendem as necessidades dos usuários (BURKE, 2002; ADOMAVICIUS; TUZHILIN, 2005).

Em outras palavras, funções de utilidade podem ser retratadas como personalizações provenientes dos usuários, ou seja, o sistema oferece ferramentas onde é possível construir uma carta dos itens que mais despertam seus interesses sendo que os mesmos selecionam valores de

atributos pré-definidos – tais como gênero de filmes, ano de lançamento, entre outros – ou descrevem com palavras – tais como nomes de músicas, cantores, entre outros – os itens desejados. Após esta etapa, o sistema faz um processo de filtragem e apresenta aqueles itens que se adéquam com as características inseridas pelos usuários. Apesar de aparentar ser um método prático e que contempla todas as necessidades de seus usuários, este apresenta algumas limitações. Ao interpelar por informações, existe a necessidade de um grande esforço do usuário para o preenchimento e quando se extrapola a muitos usuários este processo se torna ainda mais árduo.

Note que, as carências dos usuários se alteram ao longo do tempo e o sistema deve ser capaz de contemplar tais modificações para determinar novos valores de utilidade, como exemplo: Em determinada época o usuário necessita de um item específico urgentemente, sendo que o preço não é uma condição primordial e sim o prazo de entrega. Em outra, estas necessidades podem se inverter e o sistema deve ser capaz através das funções de utilidade contemplar tais casos (BURKE, 2002).

As vantagens e desvantagens desta técnica de SR é demonstrada no quadro 5 e foram obtidas a partir das pesquisas de Pazzani (1999), Burke (2002) e Schafer et al. (2007).

Quadro 8 - Vantagens e Desvantagens SR Funções de Utilidade

SR Função de Utilidade	
Vantagens	Desvantagens
Este sistema não apresenta problema de <i>Ramp-up</i> de novo usuário e novo item, justamente por não embasar suas recomendações em avaliações provenientes dos usuários.	O problema inicial para qualquer sistema com esta abordagem é desenvolver funções de utilidade que consigam exprimir corretamente as características dos itens pertencentes ao sistema.
Extremamente sensível a mudança de preferência dos usuários que utilizam deste sistema, devido o mesmo expressar seus interesses escolhendo entre diversas funções de utilidade.	Como a filtragem e consequentemente sugestões de itens são originadas a partir de quais funções de utilidade atendem as necessidades de um usuário em particular, o mesmo deve saber de forma clara quais são aquelas que realmente atendem seus anseios, caso contrário o sistema pode recomendar inúmeros itens de forma que o sistema passe a ser desnecessário.
Não inclui apenas características das funcionalidades dos itens, mas também fatores que contribuem para caracterizar as funções de estado dos mesmos.	Devido as recomendações se fundamentarem nas funções de utilidades e não nas quantidades de interações dos usuários, a habilidade de melhorar as recomendações ao longo do tempo é estática.

Fonte: Autor

4.5 SR BASEADO EM CONHECIMENTO

Assim como os SR Funções de Utilidade, SR Baseado em Conhecimento não fundamenta suas recomendações nos modelos estatísticos gerados através das avaliações provenientes dos usuários. Tal abordagem objetiva fazer recomendações baseado em métodos de inferência. As inferências são provenientes dos dados que caracterizam os itens e as preferências dos usuários. A essência desta abordagem está em inferir qual item em particular atende uma necessidade específica (que pode ser momentânea) de um usuário qualquer. Para tanto, o conhecimento em tais sistemas deve contemplar três esferas distintas, sendo elas (BURKE, 2002):

- a) conhecimento dos itens: O sistema deve ser capaz de saber quais itens estão sendo recomendados e suas características. Como exemplo pode-se de destacar um sistema que faz recomendações de ferramentas diversas, este deve ser capaz de saber que um pincel é um objeto utilizado quando a finalidade é de se pintar uma parede, muro, entre outros;
- b) conhecimento das funcionalidades: O sistema deve ser capaz de mapear as necessidades dos usuários e saber quais itens possuem as características que atendam tais carências. Exemplificando, um sistema que tem como finalidade recomendar restaurantes, uma vez que um usuário em específico deseje um jantar romântico, o mesmo deve ser capaz de sugerir um que proporcione uma refeição a luz de velas; e
- c) conhecimento dos usuários: Para que tais sistemas forneçam recomendações com boas qualidades, é necessário que o mesmo tenha formas de aprender a conhecer os usuários que os utilizam. Para tal aprendizagem dois tratamentos podem ser utilizados, sendo eles: formas de adquirir informações de cunho pessoal (semelhante a abordagem demográfica) e métodos que consigam extrair informações específicas das necessidades dos usuários.

As vantagens e desvantagens desta técnica de SR é demonstrada no quadro 6 e foram obtidas a partir das pesquisas de Pazzani (1999), Burke (2002) e Schafer et al. (2007).

Quadro 9 - Vantagens e Desvantagens SR Baseado em Conhecimento

SR Baseado em Conhecimento	
Vantagens	Desvantagens
Assim como os SR Funções de Utilidades, este método não apresenta problema de <i>Ramp-up</i> de novo usuário e novo item.	Devido as recomendações desta abordagem não se fundamentarem em métodos estatísticos que se baseiam nas avaliações provenientes dos usuários, a habilidade de melhorar as recomendações ao longo do tempo é estática.
Como tal abordagem deve ter como primícias métodos que aprendam as necessidades dos usuários de forma rápida e eficiente, este é suscetível a mudanças de necessidades de forma célere.	Para atender as três esferas que sustentam os SR Baseado em Conhecimento (Conhecimento dos Itens, Conhecimento das Funcionalidade e Conhecimento dos Usuários), são necessários algoritmos de aprendizagem de máquinas que satisfaçam as inúmeras condições que tais exigem, exemplos: formas de auferir dados, filtrá-los, processá-los e por fim apresentar respostas satisfatórias.
Semelhante também ao SR Funções de Utilidades, esta abordagem não inclui apenas características das funcionalidades dos itens, mas também fatores que contribuem para caracterizar as funções de estado dos mesmos.	
Apesar da complexidade, esta é uma das principais vantagens de tal abordagem, as necessidades dos usuários são mapeadas, portanto as recomendações apresentam altos índices de satisfação.	

Fonte: Autor

4.6 QUADRO DEMONSTRATIVO ENTRE OS MÉTODOS DE SR

Levantando os pontos importantes de cada abordagem e descrevendo-os no quadro 7, conforme as três divisões estruturais para SR, têm-se:

Quadro 10 - Características de técnicas individuais em Sistemas de Recomendação

(continua)

Abordagem	Dados Preliminares	Dados Inseridos	Processos
Baseado em Conteúdo	Características dos itens que são contemplados pelo sistema.	Avaliações dos itens provenientes dos usuários contemplados pelo sistema.	Determinar os itens que tiveram as melhores avaliações por um usuário em específico para posteriormente sugerir outros itens que compartilham das mesmas características.
Filtragem Colaborativa	Uma base de dados com as avaliações provenientes dos usuários do sistema referente a diversos itens. Note que esta base de dados é atualizada de acordo	Avaliações dos itens provenientes dos usuários contemplados pelo sistema.	Identificar e classificar em grupos distintos os usuários que compartilham das mesmas preferências. Posteriormente, sugerir para um integrante em específico de um grupo qualquer um item que o mesmo ainda não tenha avaliado, porém o grupo à qual pertença já tenha avaliado.

	com as novas avaliações provenientes dos usuários.		
--	--	--	--

Quadro 11 - Características de técnicas individuais em Sistemas de Recomendação

(conclusão)

Abordagem	Dados Preliminares	Dados Inseridos	Processos
Demográficos	Características de cunho pessoal sobre os usuários e suas avaliações referente aos itens contemplados pelo sistema. Note que esta base de dados é atualizada de acordo com os novos usuários e suas avaliações.	Características de cunho pessoal dos usuários do sistema e suas avaliações referentes aos diversos itens.	Identificar e classificar em grupos distintos os usuários que tenham características de cunho pessoal semelhantes. Posteriormente sugerir para um integrante em específico de um grupo qualquer um item que o mesmo ainda não tenha avaliado, porém o grupo à qual pertença já tenha avaliado.
Função de Utilidade	Características dos itens que são contemplados pelo sistema.	Valores para as funções de utilidade descritas pelo sistema.	Comparar os valores atribuídos as funções de utilidade com as características dos itens. Aqueles que apresentarem maior semelhança serão sugeridos aos usuários.
Conhecimento	Características dos itens contemplados pelo sistema e conhecimento de como tais itens podem atender as necessidades dos usuários contemplados pelo sistema.	Descrição dos interesses e necessidades dos usuários contemplados pelo sistema.	Inferir através de métodos de aprendizado de máquina quais itens atendem os interesses e necessidades dos usuários.

Fonte: Autor

4.7 SR HÍBRIDOS

A concepção de sistemas híbridos, teve como objetivo solucionar desvantagens apresentadas individualmente por cada abordagem. Para um sistema ser considerado híbrido, no mínimo duas abordagens de SR devem ser compreendidas. Burke (2002), classifica sistemas híbridos de acordo com a forma que são feitas as tratativas entre as abordagens. As classificações são:

- a) *weighed*: para cada abordagem presente no sistema híbrido é computado uma pontuação de acordo com as recomendações correta que o mesmo teve, sendo assim, aqueles que apresentarem melhores pontuações terão maior influência nas recomendações geradas. Como exemplo, suponha que um sistema tenha duas abordagens distintas, Filtragem Colaborativa e Baseado em Conteúdo, inicialmente

o mesmo atribui um valor de eficácia igual para ambos. Conforme os usuários vão avaliando as sugestões originadas por ambas abordagens, o sistema atribui valores maiores para aquelas que apresentam maior índice de satisfação referente as suas sugestões. Vale ressaltar que esta estratégia é relativamente simples devido apenas ponderar o peso que cada abordagem terá conforme a utilização do sistema, porém, nesta vertente de sistema híbrido é pressuposto que todas estas abordagens apresentam uma normalidade de acerto conforme o uso do sistema, o que não é uma verdade, pois algumas abordagens melhoram seu desempenho de acordo com o tempo de utilização;

- b) *switching*: esta abordagem híbrida tem como premissa fazer a alteração entre as abordagens contempladas por ele de acordo com a situação em que se encontra os elementos que compõe o SR, ou seja, itens, usuários, avaliações, funcionalidades dos itens, características de cunho pessoal e estado, entre outros. De forma prática, um sistema para evitar problemas de *ramp-up*, pode utilizar inicialmente de abordagens que não fundamentam suas recomendações em métodos estatísticos, após um determinado tempo, o sistema híbrido alterna sua abordagem para técnicas de filtragem colaborativas. Ao utilizar de tal prática de sistemas híbridos, as dificuldades individuais para cada técnica de sistema de recomendação são superadas conforme a utilização de mais de uma abordagem. Note que ao se desenvolver tais sistemas deve-se levar em conta que sua complexidade aumenta, devido o mesmo ter que atentar-se aos critérios de mudança entre as abordagens;
- c) *mixed*: todas as abordagens abrangidas pelo sistema híbrido geram recomendações. Estas recomendações são apresentadas simultaneamente ao usuário. Um exemplo prático para tal técnica é: Um sistema híbrido é desenvolvido para fazer recomendações de restaurantes de acordo com suas iguarias. O sistema compreende as abordagens funções de utilidades e filtragem colaborativa. Em relação a abordagem funções de utilidade, o usuário determina quais as características e refeições que lhe agradam, o sistema coleta esses dados e determina quais restaurantes abrangem suas preferências. Em relação a filtragem colaborativa, o sistema analisa qual grupo pertence seu usuário de acordo com suas preferências e sugere um restaurante. Por fim, as recomendações geradas em ambas as técnicas são apresentadas ao usuário;
- d) *feature combination*: é previsto a junção de dois ou mais métodos de sistemas de recomendação, onde uma abordagem é denominada principal e as outras

denominadas secundárias. As abordagens secundárias se unem a esta principal de forma que estas sejam tratadas como um aspecto a mais a ser analisado pela principal. Exemplificando esta ideia, imagine que um sistema aborde uma técnica baseada em conteúdo e filtragem colaborativa. A baseada em conteúdo é a principal e a filtragem colaborativa se une a ela como uma característica de avaliação de produtos. Todas as tratativas que envolvem o processo de recomendação proveniente da técnica baseada em conteúdo serão analisadas com a adição das avaliações provenientes da filtragem colaborativa e só assim o sistema realizará as sugestões. Note que, ao se unir duas ou mais técnicas a uma abordagem principal, estas secundárias podem deixar de influenciar com suas vantagens, isto é, para o caso apresentado anteriormente umas das vantagens da abordagem filtragem colaborativa é que a mesma apresenta sugestões de itens que não compartilham das mesmas características, quando esta passa a ser secundária esta vantagem é suprimida;

- e) *cascade*: esta técnica apresenta níveis de estágios distintos para recomendação de um item em específico. O modo mais simples tem dois estágios para elaboração da sugestão de um item qualquer. Através da interação entre usuário e sistema, a abordagem do primeiro nível faz a filtragem dos itens que atendem as necessidades, esta amostra selecionada avança para o segundo nível onde uma nova abordagem é aplicada refinando ainda mais a quantidade de itens. Suponha que um sistema híbrido contemple duas abordagens, sendo elas funções de utilidades e demográficas. No primeiro nível está embutido a abordagem de funções de utilidades. Nesta etapa um usuário em específico apresenta as características que deseja de um item específico, após conhecida as características uma série de itens são filtrados, estes vão para o segundo estágio onde através da técnica de sistemas de recomendação demográficas estes são filtrados novamente e por fim aqueles que passarem nas duas filtragens serão sugeridos;
- f) *feature augmentation*: uma abordagem inicial é utilizada para avaliação ou classificação de forma que gere novas características através de métodos de aprendizagem. Após geradas, estas características são agregadas na segunda etapa onde está uma nova abordagem de sistema de recomendação, ou seja, a saída gerada pela primeira etapa é utilizada como entrada na segunda etapa. Note que este método híbrido é muito parecido com a técnica *Cascade* e *Feature Combination*. Em relação *Feature Combination*, a diferença está em que as informações adicionadas a outras abordagens são dados crus, ou seja, não precisam de nenhum método de aprendizado

de máquina para serem determinados. Ao se comparar com *Cascade*, a segunda etapa não usa informações geradas pela primeira etapa, mas sim faz uma nova filtragem dos itens que já foram filtrados; e

- g) *meta-level*: compreende duas técnicas de recomendação. A primeira técnica é responsável por gerar um modelo sobre os usuários e/ou itens que o sistema contempla. Após gerado tal modelo, o mesmo é utilizado pela segunda técnica para gerar recomendações mais qualitativas e coerentes. Em sistemas híbridos *Meta-level* o grande benefício, principalmente para aqueles que abordam as técnicas baseado em conteúdo e filtragem colaborativa, é que os interesses dos usuários são representados através de um modelo de interesses e conseqüentemente a técnica de filtragem colaborativa que receberá tal modelo como entrada, pode tratar de forma mais adequada e facilitada de tais dados, uma vez que os mesmos já foram depurados na técnica inicial.

Como dito anteriormente, Sistemas Proativos de Recomendação podem ser desenvolvidos em diferentes grupos de aplicações com diferentes finalidades. Assim como para cada técnica individual de SR, também não existe uma técnica de SR híbrida que seja melhor que todas as outras em todos os tipos de contextos.

Vale ressaltar que estas técnicas híbridas foram mostradas e explicadas nesta seção pois muitos Sistemas Proativos de Recomendação Sensíveis ao Contexto utilizam também destas técnicas para elaborar suas recomendações. Sendo assim, ao avaliar um sistema proativo em relação a quais técnicas são empregadas pelo sistema para elaborar uma sugestão, é necessário que se tenha o conhecimento tanto das técnicas individuais como as híbridas.

Como o intuito desta dissertação não é avaliar quais técnicas híbridas são mais eficazes em diferentes contextos, no sistema proativo de recomendação de filmes desenvolvido ao final desta dissertação não é utilizado nenhuma técnica híbrida para a geração das sugestões.

4.8 ESTRUTURA SISTEMAS DE RECOMENDAÇÃO DE SPRSC

Nesta seção foi explicado quais são as características e funcionalidades para as técnicas de Sistemas de Recomendação. Conforme os dados que um Sistema Proativo de Recomendação Sensível ao Contexto possui, este pode utilizar de uma ou mais abordagens de Sistemas de Recomendação na elaboração de suas sugestões, lembrando que para cada técnica existem vantagens e desvantagens. No sistema simulado nesta dissertação e depois avaliado com o modelo de avaliação SPRSC, é utilizado da abordagem Baseado em Conteúdo, devido ao

sistema compreender diversas avaliações provenientes dos usuários e compreender também diversas características dos itens que são sugeridos pelo sistema.

Com base na literatura utilizada para o desenvolvimento desta dissertação e principalmente embasado nas pesquisas de Pazzani (1999), Burke (2002) e Schafer et al. (2007), a figura 10 demonstra a estrutura Sistemas de Recomendação do modelo de avaliação SPRSC e apresenta suas principais características. Esta estrutura compreende quais são as técnicas que podem ser aplicadas de forma individual ou conjunta para a elaboração de recomendações de itens, ações e/ou serviços.

Figura 10 - Estrutura SR de SPRSC



Fonte: Autor

5 USUÁRIOS

Nesta seção é abordado as características que definem um Usuário conhecido pelo SPRSC e um Usuário não conhecido. Também é abordado o que são os conceitos de Avaliações.

5.1 USUÁRIOS

Um usuário é definido como a pessoa ou pessoas que são o alvo das recomendações geradas por um SPRSC. Um perfil de usuário concentra diferentes tipos de dados e/ou informações, desde que estas sirvam para determinação e compreensão dos usuários dos SPRSC. Pazzani e Bilsus (2007), definiram que os dados e/ou informações mais utilizadas para determinação de perfis de usuários são divididas em duas grandes esferas:

- a) modelagem dos itens de preferência dos usuários: uma das formas para obtenção de tal modelo é através de funções probabilísticas ou estatísticas que tem como objetivo prever quais itens são de interesse dos usuários. Estes são ranqueados entre os com maior probabilidade de relevância para os menores; e
- b) históricos das interações dos usuários com SR: dados e/ou informações que retratam qualquer interação que o usuário teve com o SR e/ou itens, ações e/ou serviços contemplados pelo sistema.

Levando em conta Sistemas de Recomendação, Sistemas de Recomendação Sensíveis ao Contexto e Sistemas Proativos de Recomendação, os dados e/ou informações que caracterizam os usuários são principalmente: avaliações de preferência atribuídos aos itens, ações e/ou serviços; contexto onde estão situados e informações de cunho pessoal. Portanto, os usuários são classificados entre (BURKE, 2002; PAZZANI; BILSUS, 2007):

- a) usuários conhecidos: possui dados e/ou informações que os ajudam a ser caracterizados;
- b) usuários desconhecidos: não possui nenhum dado e/ou informação (novos usuários).

Em relação as avaliações de itens, ações e/ou serviços provenientes dos usuários, tem-se que a definição de avaliação é a representação qualitativa da satisfação de um item, ação e/ou serviço qualquer para um usuário em específico (SCHAFER et al., 2007).

As avaliações em SR, podem ser auferidas de forma explícita e implícita, sendo que as explícitas os usuários são questionados a informar qual sua opinião a respeito de um item em específico e as implícitas são determinadas através de inferências das ações dos usuários.

As avaliações explícitas podem ser simbolizadas de várias formas, destacando-se (SCHAFER et al., 2007):

- a) avaliações escalares: atribuem valores numéricos de acordo com a satisfação, como exemplo notas de 1 a 10;
- b) avaliações ordinais: atribuem um estado de identificação, como exemplos Concordo Plenamente, Concordo, Indiferente, Discordo e Discordo Plenamente;
- c) avaliações Binárias: limitam as avaliações a dois valores possíveis, como exemplo, Bom ou Ruim; e
- d) avaliações Unárias: indicam se um item foi comprado ou avaliado positivamente. Note que as ausências de tais avaliações denotam que não existe nenhuma informação proveniente do usuário referente a este item.

Em formas implícitas, os usuários e suas ações são observados, sendo que os mesmos não são interpelados por avaliações. Como exemplo, um usuário que compra um item em específico, pressupõe-se que o mesmo gostou do item, porém se este item for devolvido é um sinal de que não foi atendida suas ambições (SCHAFER et al., 2007).

Note que, ao tratar de abordagens implícitas, tais avaliações provenientes dos métodos de aprendizado possuem níveis de subjetividade altos podendo resultar em conclusões errôneas. Comparando dados obtidos através de abordagens explícitas com as adquiridas implicitamente, têm-se que as obtidas explicitamente são mais confiáveis, porém, vale destacar que itens, ações e/ou serviços avaliados de forma explícita são uma pequena porcentagem do total de itens, ações e/ou serviços que os usuários utilizam, praticam e/ou interagem (PAZZANI; BILLSUS, 2007).

5.2 ESTRUTURA USUÁRIOS DE SPRSC

Com base na literatura utilizada para o desenvolvimento deste capítulo e principalmente embasado nas pesquisas de Burke (2002), Pazzani e Billsus (2007), a figura 11 demonstra a estrutura Usuários do modelo de avaliação SPRSC e apresenta suas principais características. Este modelo compreende quais são os dois tipos de Usuários compreendidos pelos Sistemas Proativos de Recomendação e suas características.

Figura 11 - Estrutura Usuário de SPRSC



Fonte: Autor

6 APRENDIZADO DE MÁQUINA

No modelo de avaliação de SPRSC, a estrutura Aprendizado de Máquina compreende as ferramentas que permitem a análise e manipulação de dados e/ou informações afim de determinar quais recomendações gerar, para quais usuários gerar e em quais momentos.

Nesta seção, são abordados conceitos de Inteligência Artificial, paradigmas de Aprendizagem de Máquina, quais são as principais características relacionados a Aprendizagem de Máquina que devem ser avaliadas em Sistemas Proativos de Recomendação, razões e explicações dos porquês da utilização de um algoritmo em específico para a simulação de SPRSC e por fim uma sintetização de quais características são compreendidas nesta estrutura.

6.1 INTELIGÊNCIA ARTIFICIAL

Devido a Inteligência Artificial (IA) ter sua origem embasadas em diferentes áreas de conhecimento, sendo elas, Filosofia, Matemática, Economia, Psicologia, Teoria de Controle, Engenharia de Computação e Cibernética, existem diversas definições para IA (RUSSEL; NORVIG, 2004).

Como SPRSC têm como pretensão determinar itens, ações e/ou serviços para seus usuários baseado no contexto em que estão inseridos, a definição de IA utilizada nesta dissertação é a cunhada por Barr & Feigenbaum (1981) que diz: “IA é a parte da ciência da computação que se preocupa em desenvolver sistemas computacionais inteligentes, isto é, sistemas que exibem características, as quais nós associamos com a inteligência no comportamento humano – por exemplo, compreensão da linguagem, aprendizado, raciocínio, resolução de problemas, etc”.

6.2 APRENDIZADO DE MÁQUINA

Aprendizado de Máquina é uma subárea da IA e sua principal finalidade é: através de métodos computacionais adquirir novos conhecimentos e organizar os já existentes.

Simon (1983), um dos propulsores do conceito de aprendizado de máquina no âmbito da IA definiu o aprendizado da seguinte forma: “Aprendizado denota mudanças no sistema, que são adaptáveis no sentido de que elas possibilitam que o sistema faça a mesma tarefa ou tarefas sobre uma mesma população, de uma maneira mais eficiente a cada vez”.

Da elucidação acima conclui-se que o aprendizado está embasado na capacidade da adequação de um sistema a uma tarefa em específico e com o passar do tempo e através de diversas execuções desta mesma tarefa o sistema se torna mais competente.

Outra definição sobre aprendizado de máquina é dada por Tom Mitchell em seu livro *Machine Learning* de 1997, onde é afirmado que um conjunto de métodos, técnicas e ferramentas próprias para obtenção automática de conhecimento constituem o aprendizado de máquina.

Ainda segundo Mitchell:

Pode-se afirmar que um programa computacional é capaz de aprender a partir da experiência E com respeito a um grupo de tarefas T e segundo a medida de desempenho P, se seu desempenho nas tarefas T, medido segundo P, melhora com a experiência E. (MITCHELL, 1997).

Nota-se que nas definições supracitadas, alguns pontos em comum caracterizam o aprendizado de máquina, sendo eles: o aprendizado é obtido a partir da experiência na realização de uma gama de tarefas e através de novas experiências, as tarefas são realizadas de uma forma mais prática e facilitada.

Em suma, Aprendizado de Máquina permite que sistemas automatizados encontrem soluções para problemas práticos através da experiência adquirida em tarefas executadas ao longo do tempo.

Os métodos computacionais que são utilizados para a Aprendizagem de Máquinas são classificados em grupos distintos, ou seja, são separados em diferentes paradigmas, destacando-se: Modelo Simbólico, Modelo Estatístico, Modelo Baseado em Exemplo, Modelo Conexionista, Modelo Genético.

Devido esta dissertação simular um SPRSC de filmes onde o método de aprendizado de máquina escolhido foi o algoritmo de árvore de decisão C4.5 de Quinlan (1993), têm-se que este algoritmo está inserido nos Modelos Simbólicos de aprendizado de máquina, onde suas aprendizagens são geradas através de representações simbólicas de um conceito, onde tais conhecimentos são provenientes através da análise de um banco de exemplos. Por via de regra, as representações são feitas através de expressões lógicas destacando-se métodos como árvores de decisão, regras ou redes semânticas. Em outras palavras, modelos simbólicos têm como característica expressar uma hipótese de aprendizagem em uma linguagem de fácil interpretação para o usuário (SGARBI, 2007; MONARD; PRADI, 2007?).

Ainda em relação aos métodos de aprendizagem, estes são classificados também conforme as características que os mesmos possuem em relação as suas tratativas para obtenção de conhecimentos.

Tais grupos de classificação são (RUSSEL; NORVIG, 2014; SGARBI, 2007):

- a) aprendizado supervisionado: dado uma série de exemplos que apresentam valores de entrada e valores de saída, o sistema tem como objetivo determinar regras que ajudam a identificar a relação entre entrada e saída;
- b) aprendizado não-supervisionado: valores de entrada são conhecidos pelo sistema, porém suas saídas são omitidas. O objetivo dos métodos de aprendizado inseridos nesta categoria, é justamente encontrar padrões nos valores de entrada e posteriormente classificá-los em grupos distintos;
- c) aprendizado semi-supervisionado: tal categoria mescla as duas supracitadas. Usualmente, apenas alguns valores de entrada apresentam valores de saída e um vasto conjunto de entrada não apresenta valores de saída, cabendo ao sistema classificá-los; e
- d) aprendizado por reforço: um sistema aprende a agir em um ambiente dinâmico afim de atingir um objetivo específico. Para tanto, a cada ação que o sistema executa, o ambiente lhe fornece um valor de recompensa, podendo ser positiva ou negativa, que servirá como método de aprendizado.

Dentre de tais categorias supracitadas, os métodos de aprendizagem podem ser dispostos em duas classes distintas. Estas classes se diferenciam em relação a forma com que é tratado os exemplos que são utilizados para o aprendizado, sendo elas (SGARBI, 2007):

- a) não-incremental: é necessário que todos os exemplos para aprendizagem estejam simultaneamente disponíveis; e
- b) incremental: a hipótese gerada é atualizada de forma corrente, sempre que novos exemplos são adicionados ao conjunto de treinamento.

Devido os dados utilizados para a simulação do SPRSC desta dissertação serem classificados e já estarem armazenados em suas respectivas bases de dados, o algoritmo escolhido para aprendizagem de máquina, C4.5 de Quinlan (1993), está inserido nos Modelos de algoritmo Simbólicos que utilizam de Aprendizagem Supervisionada Não-Incremental.

6.3 APRENDIZAGEM INDUTIVA

Graças a IoT, dados heterogêneos provenientes de diversos sensores e em grandes quantidades são adquiridos afim de caracterizar diversos contextos. Em SPRSC, diferentes bases de dados são compreendidas sendo que ao aplicar métodos de Aprendizado de Máquina nestes bancos, é possível determinar quais itens, ações e/ou serviços recomendar e em quais momentos fazer estas recomendações. Como grande parte destes bancos de dados possuem valores de entrada e saída, métodos de Aprendizado de Máquina Supervisionados podem ser aplicados (PERERA et al., 2014; SALMAN et al., 2015).

Métodos de Aprendizado de Máquina Supervisionado tem suas fundamentações no conceito de Aprendizagem Indutiva. Hume em 1739, definiu o princípio da indução como a obtenção de regras gerais através de exposições a associações repetidas. Em aprendizado de máquina, Mitchell (1997) definiu Aprendizagem Indutiva da seguinte forma: Dado um par $(x, f(x))$, onde x é a variável de entrada e $f(x)$ é o valor da saída da função quando aplicado x , a tarefa da inferência indutiva pura (ou indução) é retornar uma função h (hipótese) que se aproxime a função f dado uma coleção de exemplos de f (MITCHELL, 1997; RUSSEL; NORVIG, 2014).

Em algumas circunstâncias, exemplos de pares $(x, f(x))$ podem ser descritas de formas consistentes por várias hipóteses, gerando a indagação de qual hipótese consistente escolher já que todas apresentam resultados coerentes. Para tal questão é utilizada a lâmina de Ockham que diz: “prefira a hipótese mais simples consistente com os dados”. Tal definição faz sentido, pois as hipóteses que não são mais simples que os dados deixam de extrair algum padrão dos mesmos (MITCHELL, 1997; RUSSEL; NORVIG, 2014).

A forma com a qual se mensura o desempenho de um método de aprendizagem é através da análise da classificação de exemplos não contidos no conjunto de treinamento. As decisões apresentadas que são plausíveis com a realidade é um bom indicativo para demonstrar que a técnica está coerente com a situação em que foi empregada, já aquelas que não são plausíveis são indicativos de que o método de aprendizado não é bom para a situação em que está inserido. Mitchell (1997) apresenta uma metodologia para analisar o desempenho de um método específico, sendo ele:

- a) coletar um grande conjunto de exemplos;
- b) dividi-los em dois conjuntos disjuntos: conjunto de treinamento e conjunto de teste;
- c) aplicar o método de aprendizagem ao conjunto de treinamento, gerando uma hipótese h ;

- d) medir a porcentagem de exemplos no conjunto de teste que são corretamente classificados por h; e
- e) repetir as etapas 1 a 4 para diferentes conjuntos de treinamento selecionando-os aleatoriamente.

6.4 APRENDIZAGEM DE MÁQUINA EM SPR

Após lido a literatura referente a diversos artigos sobre Sistemas de Recomendação Sensíveis ao Contexto e Sistemas Proativos de Recomendação, algumas características foram determinadas e distribuídas nesta estrutura do modelo de avaliação de Sistemas Proativos de Recomendação.

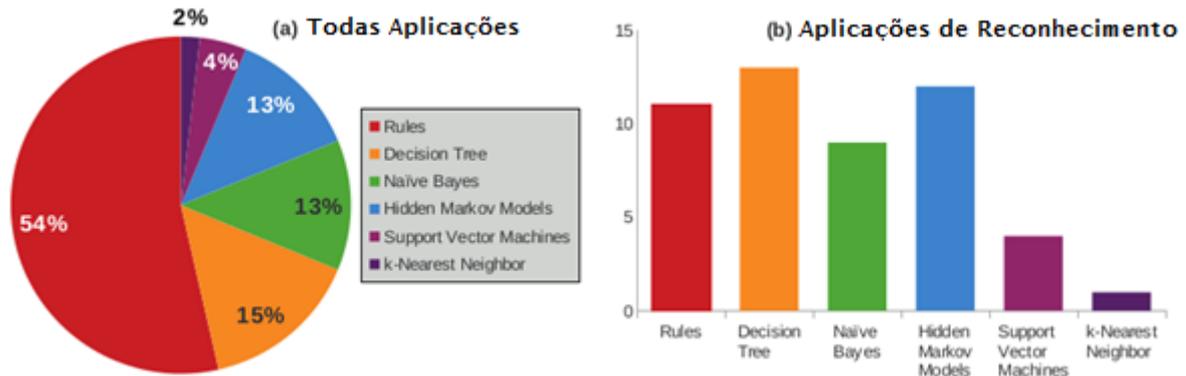
- a) a primeira é em relação aos métodos de Aprendizagem do sistema. Neste item é declarado qual algoritmo é utilizado para que o sistema aprenda a determinar as preferências dos usuários e os momentos em que um usuário necessita de uma recomendação proativa baseado no contexto onde estão inseridos. Com base nas informações em que os SPR têm do usuário e do contexto onde estão inseridos, o sistema proativo, através do Aprendizado de Máquina compreendido por ele, determina quais recomendações fazer e em quais momentos (MUNOZ-ORGANERO et al., 2010; PERERA et al., 2014);
- b) o segundo item, está relacionado as regras que descrevem o contexto onde está inserido o Sistema Proativo de Recomendação e como estas regras são utilizadas para determinar se o sistema deve ou não fazer uma sugestão. Exemplificando esta ideia, imagine um SPR que faça sugestões de filmes para seus usuários. Dado que um usuário seja uma criança, este sistema deve se atentar as regras relativas as classificações dos filmes, ou seja, a faixa etária que determina se crianças podem ou não podem assistir ao filme. Portanto, mesmo que o filme tenha as características que agradem esta criança e o momento seja pertinente para uma sugestão, caso este filme não atenda a regra de classificação ele não deve ser recomendado (GALLEGO; WOERNDL; HUECAS, 2013); e
- c) o terceiro item avalia quais métricas são utilizadas para determinação de valores de limiar. Lembrando que valores de limiar determinam em quais momentos os SPR devem fornecer uma sugestão autônoma, ou seja, sem que o usuário peça por uma sugestão explicitamente. Muitas vezes o valor de limiar é determinado através da relação entre as preferências do usuário, momentos de pertinência e as regras que

descrevem o contexto onde o sistema e usuário estão inseridos (GALLEGO; WOERNDL; HUECAS, 2013; SALMAN, 2015).

6.5 APRENDIZAGEM INDUTIVA ATRAVÉS DE ÁRVORES DE DECISÃO

Para aferição de Contexto e elaboração de diferentes sistemas que geram recomendações uma série de métodos de Aprendizado de Máquina são utilizados. A figura 12a, demonstra quais são as porcentagens dos métodos de aprendizado utilizados em 109 aplicações que levam em consideração o contexto onde está o sistema. A figura 12b, demonstra quais métodos são mais utilizados para classificação e reconhecimento de situações em 50 aplicações.

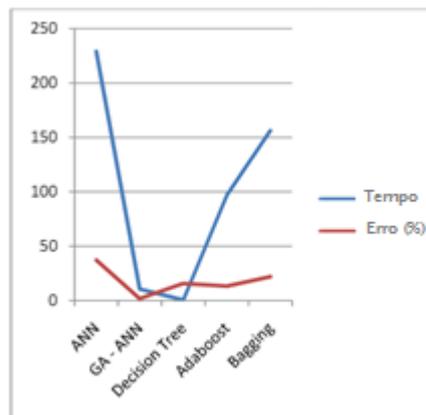
Figura 12 - Métodos de Aprendizado de Máquina Utilizados



Fonte: Autor “adaptado de” Perera et al., 2014

Na figura 13, está demonstrado uma análise sobre o tempo que leva o aprendizado para cada técnica de Aprendizado de Máquina e sua precisão.

Figura 13 - Tempo vs Erro de técnicas de Aprendizado de Máquina em SR na IoT



Fonte: Autor “adaptado de” Sachdeva, Dhir e Kumar, 2016

Devido a IoT fornecer grandes quantidades de dados classificados e os resultados apresentados nas figuras acima, para a simulação do sistema de recomendação compreendido nesta dissertação e posteriormente avaliado através do modelo de avaliação de Sistemas Proativos de Recomendação desenvolvida nesta dissertação, é utilizado o método de Aprendizado de Máquina Árvores de Decisão (*Decision Trees*).

Árvores de Decisão, tem como entrada um objeto ou situação retratados por uma série de atributos e através da análise destes atributos retorna uma decisão. Tais entradas podem ser discretas ou contínuas. O valor de saída (denominada classe) também pode ser classificado como discreta ou contínua. Como dito anteriormente, quando a decisão retorna um valor discreto a aprendizagem é chamada de classificação, caso contrário, retorna uma decisão com valor contínuo, esta é chamada de aprendizagem de regressão (RUSSEL; NORVIG, 2014).

A indução nas árvores de decisão a partir de exemplos é constituída por um vetor de atributos de entrada e um único valor de saída. O termo que é designado para um conjunto completo de exemplos é conjunto de treinamento. A árvore por si só, apenas memoriza as observações contidas no conjunto de treinamento, sendo incapaz de extrapolar decisões para eventos que ainda não contempla. Para que a árvore aprenda a tomar decisões de exemplos que não existam no conjunto de treinamento, através de heurísticas encontra-se o atributo que tem mais impacto para a classificação de um exemplo. Tais atributos mais impactantes, na hora da decisão são alocados em uma árvore de decisão seguindo uma ordem de classificação do mais importante para o menor (RUSSEL; NORVIG, 2014).

O método Árvores de Decisão para aprendizado de máquinas é aconselhável para problemas que possuem algumas características em especiais, sendo elas (MITCHELL, 1997):

- a) instâncias (objetos que representam uma situação) são retratados por atributos com valores fixados: instâncias são caracterizadas por atributos fixados (por exemplo “Temperatura”) e seus valores. A forma mais simples para o aprendizado, utilizando a técnica de Árvores de Decisão, se dá quando os atributos são representados por uma pequena quantidade de valores pré-determinados (como exemplo para a instância “Temperatura” tem-se os valores pré-determinados “Quente”, “Moderado” e “Frio”). Existem extensões deste tipo de aprendizado que compreendem valores com números reais para retratar um atributo;
- b) as funções de saída possuem valores discretos: Árvore de Decisão tem como saída valores booleanos (sim ou não) ou valores que assumem mais de duas saídas possíveis;

- c) descrição de lógicas disjuntivas são exigidas na aprendizagem: naturalmente o aprendizado através de Árvore de Decisão retornam funções lógicas disjuntivas;
- d) conjunto de treinamento pode contemplar erros: este aprendizado é aos erros provenientes do conjunto de treinamento; e
- e) conjunto de treinamento tem valores para atributos não determinados: este método pode ser empregado inclusive quando no conjunto de treinamento alguns atributos não estão caracterizados ou valorados.

As bases de dados em SPRSC, compreendem dados e/ou informações provenientes de sensores inseridos na IoT. Lembrando que a IoT fornece dados de diversos tipos e em grandes quantidades, portanto é necessário que uma análise seja feita afim de filtrá-los. Estes dados podem ser contínuos, discretos e caso algum problema aconteça na aferição, estes dados podem ficar desconhecidos. Sendo assim, é necessário que um algoritmo de Árvore de Decisão compreenda estas situações e as contorne.

Quinlan (1993) desenvolveu o algoritmo C4.5 e tem tratativas para as situações acima. As características que o descreve são:

- a) atributos com valores contínuos são permitidos na base de treinamento e estes são classificados conforme um valor de limiar que o próprio algoritmo calcula;
- b) valores desconhecidos são representados com o símbolo “?”, e as tratativas para validação do Ganho de Informação sofrem alterações, ou seja, a relação do atributo de entrada que contém estes valores desconhecidos com o atributo de saída sofre um decréscimo em suas relações;
- c) ao invés de valorar a representatividade de um atributo de entrada com o atributo de saída apenas conforme a fórmula de Ganho da Informação, este usa também uma fórmula denominada Proporção de Ganho. Sendo que tal método proporciona resultados de representatividade mais confiáveis;
- d) uma vez gerada a árvore de decisão, tal algoritmo apresenta métodos para poda de árvore, ou seja, cálculos estatísticos e/ou probabilísticos que permitem a simplificação das árvores geradas; e
- e) regras são geradas a partir das árvores de decisão determinadas, sendo que tais regras são analisadas e se possível simplificadas também.

Vale ressaltar que o algoritmo C4.5 é um dos mais utilizados na literatura, devido mostrar ótimos resultados em problemas de classificação e apresentar bons desempenhos em relação a altas taxas de processamento de dados (QUINLAN, 1993).

C4.5 possui código-fonte disponível e seu Pseudocódigo é apresentado na figura 14. Uma ferramenta de análise de dados denominada Weka (utilizada nesta dissertação), utiliza de uma implementação denominada J48 que é um código aberto em Java do algoritmo C4.5.

Figura 14 - Pseudocódigo C4.5

```

Função C4.5
(R: conjunto de atributos não classificadores,
C: atributo classificador,
S: conjunto de treinamento) devolve uma árvore de decisão;

Início
Se S está vazio,
    devolver um único nó com Valor Falha;
Se todos os registros de S têm o mesmo valor para o atributo classificador,
    devolver um único nó com tal valor;
Se R está vazio, então
    devolver um único nó com o valor mais freqüente do atributo
    classificador nos registros de S [Nota: existirão erros, isto é,
    registros que não estarão bem classificados neste caso];
Se R não está vazio, então
    D-atributo com maior Proporção de Ganho(D,S) entre os atributos de R;
    Sejam {dj | j=1,2, .., m} os valores do atributo D;
    Sejam {Sj | j=1,2, .., m} os subconjuntos de S correspondentes aos
    valores de dj respectivamente;
    devolver uma árvore com a raiz nomeada como D e com os arcos nomeados
    d1, d2, .., dm que vão respectivamente às árvores
    C4.5(R-{D}, C, S1), C4.5(R-{D}, C, S2), .., C4.5(R-{D}, C, Sm);

Fim

```

Fonte: Sgarbi, 2007

O primeiro passo a ser dado no algoritmo C4.5 é calcular o valor de entropia da instância de saída de uma base de dados. O termo entropia em árvores de decisão se refere a medida da incerteza (aleatoriedade) de um evento ocorrer. A fórmula 1 demonstra como é feito tal cálculo (QUINLAN, 1993; MITCHELL, 1997).

$$Entropia(S) = - \sum_{i=1}^c p_i \log_2 p_i \quad (1)$$

O segundo passo, é calcular a medida de efetividade de um atributo (Ganho da Informação) em relação a um atributo específico em análise inserido na base de treinamento. Note que, um único atributo pode fornecer apenas uma parte da entropia calculada acima, portanto a medição deve ser fundamentada na entropia remanescente após o teste deste atributo. A fórmula 2 representa tal conceito.

$$Ganho(S, A) = Entropia(S) - \sum_{v \in \text{Valores}(A)} \frac{|S_v|}{|S|} Entropia(S_v) \quad (2)$$

O terceiro passo é calcular a Proporção de Ganho dos atributos e é dado pela seguinte fórmula 3.

$$\text{Propor. De Ganho } (A) = \frac{\text{Ganho } (S, A)}{\text{Entropia_Atributo } (A)} \quad (3)$$

O atributo que apresentar o maior valor de Proporção de Ganho será escolhido como nó raiz da árvore de decisão. De forma recursiva, os outros atributos serão relacionados ao atributo com maior valor de Proporção de Ganho e novamente serão executados os cálculos de entropia, ganho da informação e proporção de ganho (QUINLAN, 1993).

Como dito anteriormente, o C4.5 compreende valores contínuos na base de treinamento. Definindo formalmente como tais valores são tratados, tem-se que: supondo um atributo contínuo X , mesmo que seu domínio possa ter infinitos valores, os números de exemplos em um conjunto de treinamento T é finito, sendo assim, o número de valores diferentes para esse atributo também é finito (QUINLAN, 1993).

Ordenando de forma crescente os valores contínuos do atributo $X \{a_1, a_2, \dots, a_m\}$, onde T apresente dois valores de classificação T_1 e T_2 , tem-se que alguns valores do atributo X estarão no conjunto T_1 e outros no conjunto T_2 .

Uma vez ordenado os valores do atributo X , sempre que a classificação se alterar entre os conjuntos T_1 e T_2 , é calculado através da fórmula 4 o valor médio entre os valores, resultando em um intervalo específico. Para cada intervalo é calculado o ganho e/ou ganho proporcional e aquele que apresentar o maior valor será considerado o nó de decisão.

$$\text{Valor_Medio} = \frac{a_{(i) \in T_1} + a_{(i+1) \in T_2}}{2} \quad (4)$$

Exemplificando tal conceito, será utilizado o conjunto de treinamento T abaixo, onde em T_1 estão os valores que são classificados como Sim e T_2 estão os valores que são classificados como Não.

Figura 15 - Base de Treinamento com Valores Contínuos

Elemento	Atributo	Classe	
a_1	64	Sim	
a_2	65	Não	→ $(64 + 65) / 2 = 64,5$
a_3	68	Sim	→ $(65 + 68) / 2 = 66,5$
a_4	69	Sim	
a_5	70	Sim	
a_6	71	Não	→ $(70 + 71) / 2 = 70,5$
a_7	72	Não	→ $(72 + 73) / 2 = 72,5$
a_8	73	Sim	
a_9	75	Sim	
a_{10}	75	Sim	→ $(75 + 80) / 2 = 77,5$
a_{11}	80	Não	→ $(80 + 81) / 2 = 80,5$
a_{12}	81	Sim	
a_{13}	83	Sim	→ $(83 + 85) / 2 = 84$
a_{14}	85	Não	

Fonte: Autor

Note que existem 7 intervalos onde são alterados os valores de classe. Para determinação de qual intervalo utilizar como um nó de decisão, é necessário calcular o Ganho de Informação e/ou Proporção de Ganho para cada variação e determinar qual apresenta maior valor.

Em relação a poda de árvore, o C4.5 permite a diminuição da quantidade de nós em uma árvore de decisão resultando em critérios simplificados. Existem duas classes em que os processos de poda de árvore são classificados, sendo elas (WITTEN et al., 2016):

- a) *stopping* ou *prepruning*: ao longo da geração de uma árvore de decisão, são utilizados critérios de avaliação de relevância referente a criação ou não criação de um novo nó. Caso estes critérios de avaliação não sejam atingidos, o nó que seria criado recebe a forma de folha com o valor mais pertinente de uma decisão em específico. Como exemplo destes critérios de avaliação, têm-se cálculos estatísticos, reduções de erro, ganho da informação, entre outros; e
- b) *post-pruning*. A geração da árvore é executada até o final com intuito de encontrar todas as possíveis interações de uma decisão dado um atributo em específico. Após a árvore gerada, serão analisados os nós (onde diferentes métodos de análise podem ser usados) e estes podem ser substituídos por folhas de decisão ou removidos da árvore.

No algoritmo C4.5 é utilizado um método de poda *Post-Pruning* baseado na estimativa pessimista do erro, ou seja, é avaliado a taxa de erro para árvore e seus componentes a partir do próprio conjunto de treinamento e posteriormente tais erros são comparados a esta estimativa pessimista. A estimativa pessimista do erro em um nó ou folha em específico é dada pela fórmula 5 (QUINLAN, 1993; WITTEN et al., 2016).

$$e = \frac{f + \frac{z^2}{2N} + z \sqrt{\frac{f}{N} - \frac{f^2}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}} \quad (5)$$

No algoritmo C4.5 a poda é feita por uma análise *bottom-up*, ou seja, as primeiras folhas a serem analisadas são aquelas do último nível de uma ramificação da árvore. Conforme as podas são feitas, novas análises são executadas.

Valores desconhecidos são representados com “?”. O valor da entropia do atributo de saída deve ser calculado sem levar em conta a situação onde se tem um valor desconhecido. Após calculada a entropia do atributo de saída é calculado de forma ponderada os valores de entropia dos atributos de entradas relacionados com o atributo de saída (QUINLAN, 1993).

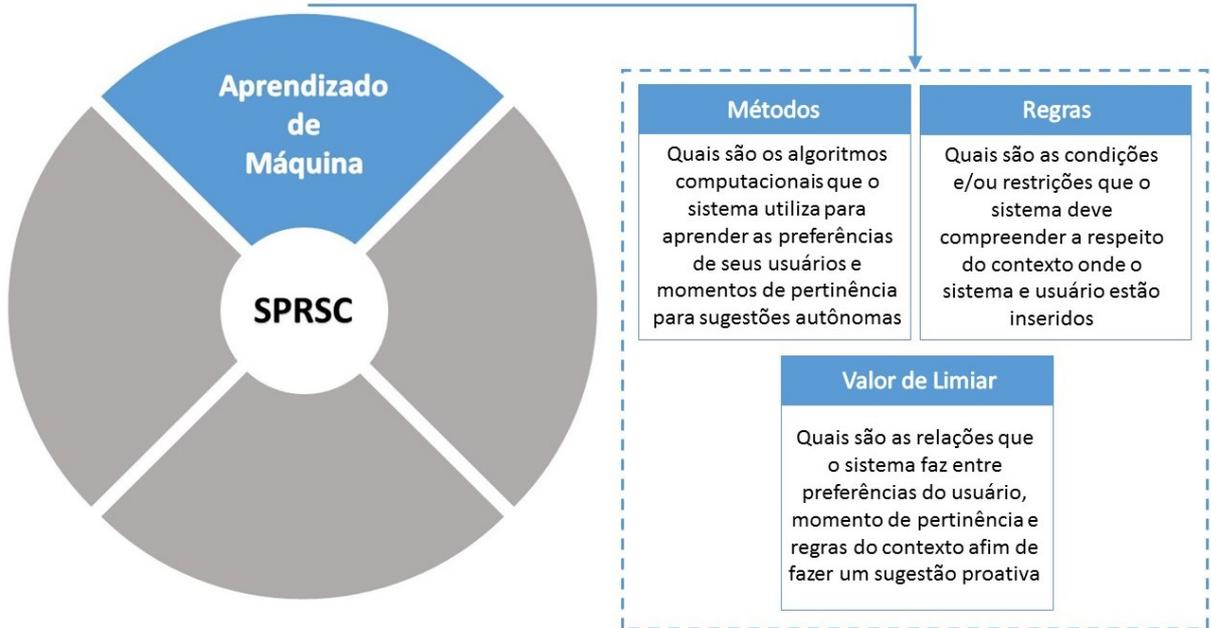
Regras de Indução são determinadas a partir de uma árvore de decisão gerada. Para tanto, ao se elaborar uma regra o nó pai deve ser analisado até uma folha de decisão. Para cada folha existe uma regra em específico. Os valores dos atributos presentes do nó pai até a folha de decisão devem ser compreendidos na regra (QUINLAN, 1993; WITTEN et al., 2016).

Estas regras determinam quais são as principais características para a tomada de decisão. Transportando este conceito para um SPRSC que faz recomendações de filmes, estas regras retratariam quais são as principais características que um usuário deseja que um filme tenha.

6.6 ESTRUTURA APRENDIZADO DE MÁQUINA DE SPRSC

A figura 16 demonstra a estrutura de Aprendizado de Máquina do modelo de avaliação SPRSC de acordo com a literatura estudada a respeito de Sistemas Proativos de Recomendação e Sistemas de Recomendação Sensíveis ao Contexto. Conforme a subseção 6.4, os itens são separados e caracterizados de acordo com suas funções e características.

Figura 16 - Estrutura Aprendizado de Máquina em SPRSC



Fonte: Autor

7 MODELO DE AVALIAÇÃO ESTRUTURADO DE SISTEMAS PROATIVOS DE RECOMENDAÇÃO SENSÍVEIS AO CONTEXTO E SIMULAÇÃO

Após cada estrutura do modelo de avaliação de SPRSC ter sido explicado, a figura 17 mostra o modelo com todas as características de todas as estruturas. Este modelo tem como pretensão avaliar os aspectos mais importantes dos Sistemas de Recomendação que também utilizam de dados do Contexto para elaborar e determinar momentos propícios para fornecer suas sugestões de forma autônoma (proativa).

Nesta seção é apresentada também o desenvolvimento de um Sistema Proativo de Recomendação Sensível ao Contexto que faz sugestões de filmes. Este sistema identifica momentos pertinentes para fazer sugestões de filmes e quais são os filmes que atendem as preferências individuais de seus usuários. A metodologia de desenvolvimento é explicada detalhadamente nesta seção e as características e resultados são discutidos com base nas estruturas do modelo de avaliação da figura 17.

7.1 BASES DE DADOS DO SPRSC DE FILMES

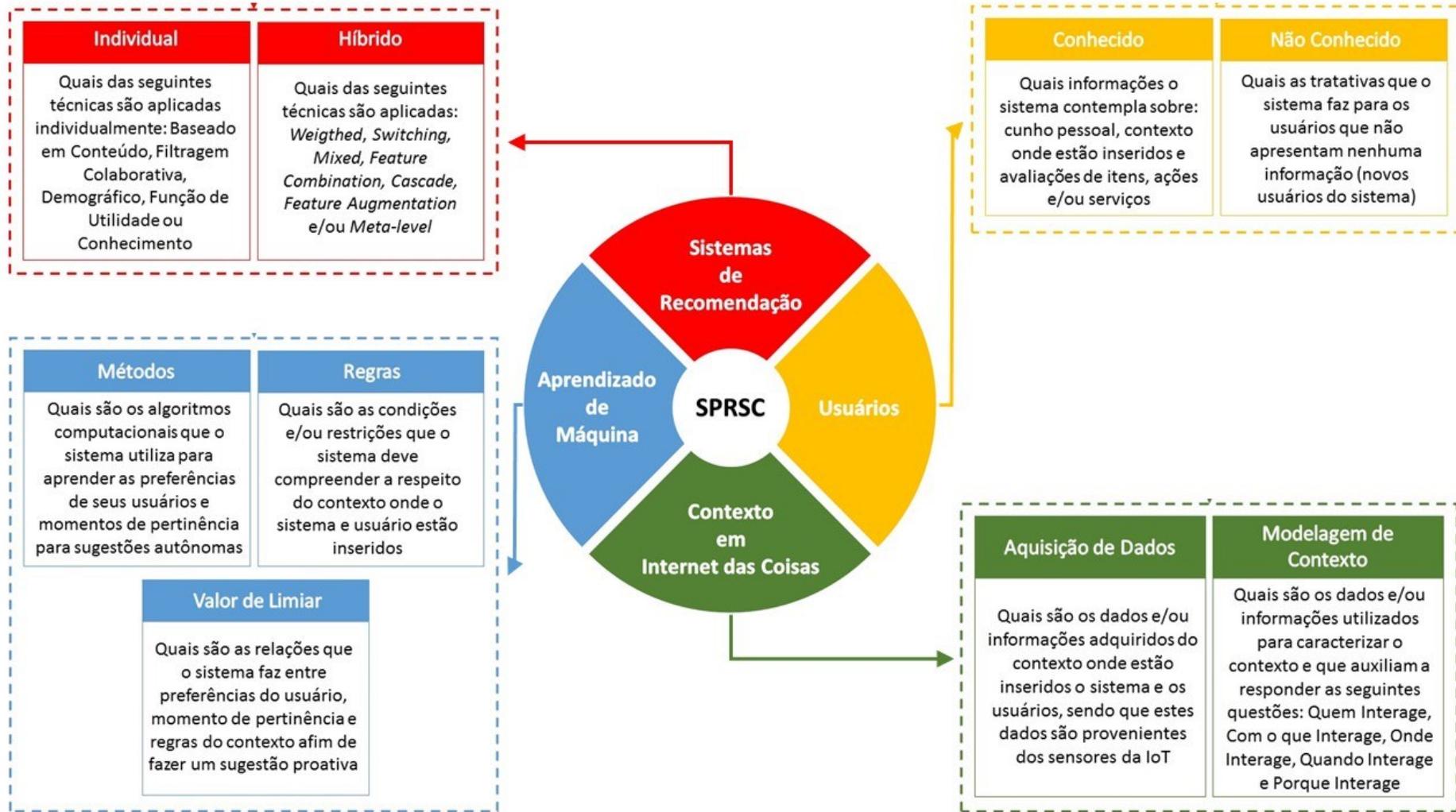
O Sistema Proativo de Recomendação Sensível ao Contexto que faz sugestões de filmes de forma autônoma (proativa) desta dissertação, foi desenvolvido a partir de duas bases de dados principais. De forma sucinta, a primeira base contém dados sobre quais momentos dentro de uma sala IoT são propícios para se assistir um filme. A segunda base contém diversas avaliações de filmes de diversos usuários.

7.1.1 Base de dados sala IoT

Esta base de dados foi criada com o objetivo de servir de fonte de dados para aplicação do algoritmo C4.5 de Quinlan (1993). A partir deste algoritmo, árvores de decisão foram geradas e através destas árvores são determinados quais momentos são pertinentes para que os usuários assistam a filmes dentro desta sala.

De acordo com o explicado e detalhado na seção Contexto em Internet das Coisas (seção 3), um ambiente IoT é caracterizado pela presença de diversos tipos de sensores que consigam se comunicar e gerar dados. Estes dados quando analisados individualmente e/ou conjuntamente permitem a modelagem do contexto onde os sensores estão inseridos. Sendo assim, para o desenvolvimento deste SPRSC, uma sala IoT hipotética contendo diversos

Figura 17 – Modelo de Avaliação de SPRSC



Fonte: Autor

sensores foi simulada através de perguntas a diferentes pessoas. Esta sala hipotética compreende diversos sensores que são capazes de:

- a) período da semana: este sensor consegue aferir se é final de semana ou dia da semana. Transpondo para um cenário real, estes dados podem ser determinados através de qualquer dispositivo que contenha um calendário e esteja sincronizado com os dias condizentes da Cidade onde esta sala IoT está localizada
- b) companhia: este sensor consegue aferir se o usuário está sozinho ou acompanhado. Transpondo para um cenário real, estes dados podem ser determinados através de etiquetas de identificação RFID, permitindo que a sala IoT, através destas etiquetas, consiga contabilizar a quantidade de pessoas presentes;
- c) estado do sofá: este sensor consegue aferir se o usuário está sentado ou não no sofá. Transpondo para um cenário real, estes dados podem ser determinados através de sensores piezoelétricos instalados nos assentos do sofá;
- d) período do dia: este sensor consegue aferir se é manhã, tarde, noite ou madrugada. Transpondo para um cenário real, estes dados podem ser determinados através de qualquer dispositivo que contenha as horas e esteja sincronizado com o horário da cidade onde a sala IoT está localizada;
- e) luminosidade: este sensor consegue aferir se a sala está iluminada ou escura. Transpondo para um cenário real, estes dados podem ser determinados através de sensores de iluminação;
- f) temperatura: este sensor consegue aferir se está frio, moderado ou quente. Transpondo para um cenário real, estes dados podem ser determinados através de sensores de temperatura; e
- g) estado da televisão: este sensor consegue determinar se a televisão está ligada ou desligada. Transpondo para um cenário real, estes dados podem ser determinados através de algoritmos de visão computacional que conseguem determinar se a tv está ligada ou desligada.

Para determinação desta base, uma pesquisa compreendendo valores aleatórios para os sensores listados acima (diferentes contextos) foi realizada com 20 pessoas de diferentes profissões e faixas etárias. Esta pesquisa indagava se elas assistiriam um filme de acordo com o contexto em questão. Cada pessoa respondeu a 25 exemplos de contextos

aleatórios, sendo que estas perguntas poderiam se repetir. No total foram obtidas 500 respostas que diziam se o momento era propício ou não para se assistir um filme. Tal base de dados está disponível no repositório github.com/Nahmir/SPRSC.

7.1.2 Base de dados avaliações de filmes

A segunda base de dados foi retirada do site grouplens.org/datasets/movielens onde diversas pessoas assistiram e avaliaram diversos filmes de acordo com seus gostos. Os atributos que caracterizavam os filmes nesta base de dados inicial eram: Nome do Filme, Ano de Lançamento e Gênero do Filme. A partir desta base inicial, alguns dados foram tratados e outros atributos em relação as características dos filmes foram adicionados, resultando nos seguintes atributos:

- a) ID de usuário: atributo que identifica qual é o usuário que fez uma avaliação;
- b) título: nome do filme;
- c) ano de lançamento: qual foi o ano de lançamento do filme;
- d) gênero 1: qual é o principal gênero do filme;
- e) gênero 2: qual é o gênero secundário do filme;
- f) classificação: qual é a faixa etária indicada para o filme;
- g) duração: quantos minutos o filme tem de duração;
- h) diretor: nome do diretor principal do filme;
- i) diretor ganhou Oscar: caso o diretor tenha sido premiado com o troféu Oscar este atributo tem o valor Sim, caso não tenha sido premiado tem o valor Não;
- j) ator principal: nome do ator principal do filme;
- k) ator principal ganhou Oscar: caso o ator tenha sido premiado com o troféu Oscar este atributo tem o valor Sim, caso não tenha sido premiado tem o valor Não;
- l) país de origem: qual foi a nacionalidade do estúdio que produziu o filme;
- m) orçamento: quanto custou para o filme ser feito;
- n) filme ganhou Oscar: caso o filme tenha sido premiado com o troféu Oscar em qualquer categoria este atributo tem o valor Sim, caso não tenha sido premiado tem o valor Não;
- o) nota IMDB: avaliação escalar atribuída ao filme pelos usuários do site IMDB (o mais relevante em avaliações de filmes); e

p) avaliação: este atributo caracteriza se o usuário gostou ou não do filme. Caso tenha gostado o atributo tem valor Sim, caso não tenha gostado tem o valor Não.

Alguns atributos desta base de dados foram utilizados para a determinação de preferências individuais dos usuários do SPRSC. Para tanto, foi utilizado também o algoritmo C4.5. Este algoritmo determinava quais características dos filmes tinham maior relevância nos gostos dos usuários do SPRSC.

Lembrando que esta base de dados está disponível também no repositório github.com/Nahmir/SPRSC.

7.2 APRENDIZADO DE MÁQUINA

Conforme dito anteriormente, foi definido para este sistema a utilização do algoritmo supervisionado de Árvore de Decisão C4.5 de Quinlan (1993), lembrando que este algoritmo foi explicado na subseção 6.5 da seção Aprendizado de Máquina. Este algoritmo foi escolhido pois trabalha com valores discretos e contínuos, faz poda de árvore, trabalha com valores desconhecidos e gera regras de indução. Para aplicação deste algoritmo, foi utilizado a ferramenta de aprendizado de máquina denominado WEKA onde o algoritmo similar ao C4.5 é denominado J48.

O algoritmo para gerar as árvores de decisão foi aplicado nas duas bases de dados do SPRSC que faz sugestões de filmes. A metodologia de aplicação e os resultados serão explicados a seguir.

7.2.1 C4.5 aplicado na base de dados sala IoT

Ao aplicar o C4.5 nesta base, foram determinadas árvores de decisão que classificavam se o momento era pertinente ou não para se assistir filmes. Estes momentos se diferenciavam de acordo com as características do contexto, ou seja, de acordo com os valores dos sensores inseridos nesta sala hipotética. Note que a sala IoT hipotética compreende as classificações de diversos usuários, portanto a árvore de decisão gerada representa a sala IoT e não apenas as preferências individuais de um usuário em específico.

Para aplicação do C4.5 foram utilizados os seguintes atributos de entrada: Período da Semana, Companhia, Estado do Sofá, Período do Dia, Luminosidade, Temperatura e

Estado da Televisão. Como atributo de saída foi utilizado a pergunta de classificação se o usuário Assistiria um Filme.

Para treinamento e posteriormente teste das árvores de decisão geradas, esta base de dados foi separada em dois conjuntos, sendo eles:

- a) conjunto de treinamento: o valor máximo de instâncias é de 400, sendo que o algoritmo C4.5 foi aplicado a diversos valores de instâncias;
- b) conjunto de teste: este conjunto é constituído por 75 instâncias classificadas. Este valor representa aproximadamente 20% do valor máximo de instâncias do conjunto de treinamento.

Após separado os conjuntos, o algoritmo C4.5 foi aplicado no conjunto de treinamento onde os valores de instâncias classificadas eram variados. De acordo com estas variações, distintas árvores de decisões eram obtidas. A tabela 1 demonstra os resultados obtidos. Note que, para cada linha onde o valor do número de instâncias é diferente é gerado uma Árvore de Decisão. Para cada árvore, as mesmas 75 instâncias do conjunto de teste são aplicadas.

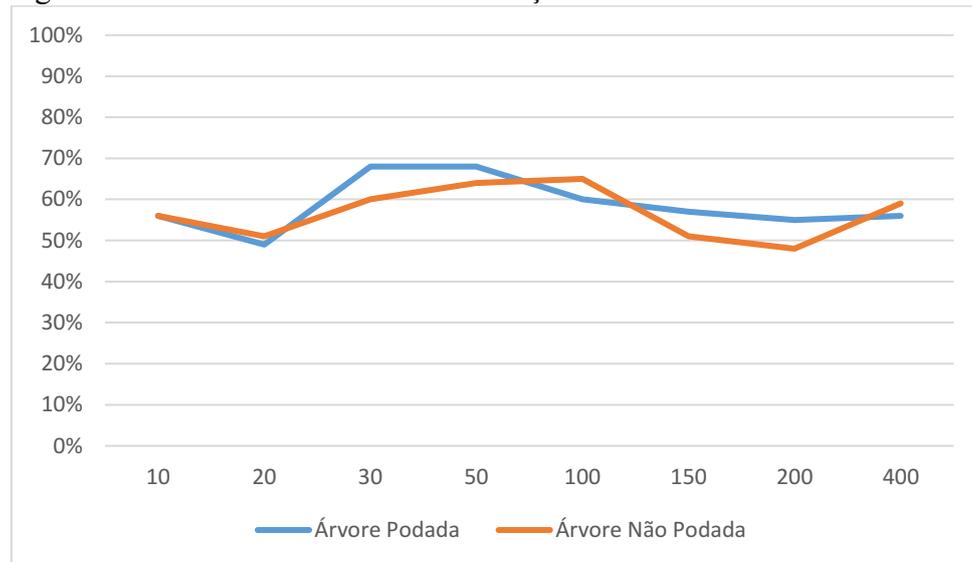
Tabela 1 - Resultados das Árvores de Decisão gerada para a sala IoT hipotética

Número de Instâncias	Acertos com Árvore Podada	Acertos com Árvore Não Podada
10	56 %	56 %
20	49 %	51 %
30	68 %	60 %
50	68 %	64 %
100	60 %	65 %
150	57 %	51 %
200	55 %	48 %
400	56 %	59 %

Fonte: Autor

Conforme os resultados obtidos na tabela 1, a figura 18 mostra a curva de porcentagem de acertos de acordo com as quantidades de instâncias utilizadas para gerar as árvores de decisão. Como o intuito desta dissertação não é discutir parâmetros de eficácia e velocidade da aplicação de algoritmos, os resultados obtidos não são comparados a nenhum outro método de aprendizado de máquina.

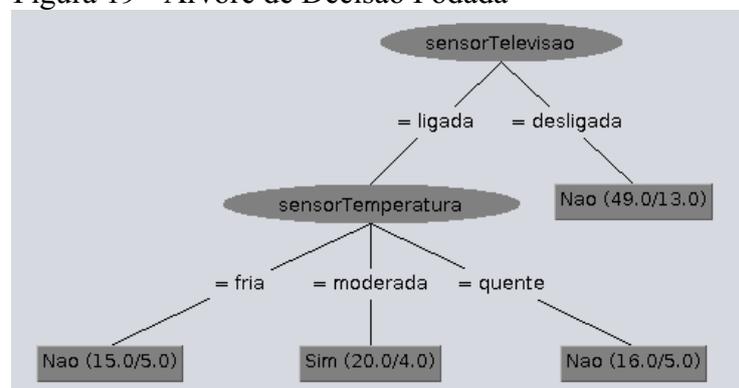
Figura 18 - Acertos de acordo com variação de Instâncias



Fonte: Autor

Para exemplificar uma árvore de decisão gerada através do algoritmo C4.5, a figura 19 mostra uma árvore de decisão podada gerada a partir de 50 instâncias classificadas do conjunto de treinamento. As outras árvores geradas estão disponíveis e podem ser visualizadas no repositório github.com/Nahmir/SPRSC.

Figura 19 - Árvore de Decisão Podada



Fonte: Autor

7.2.2 C4.5 aplicado na base de dados avaliações de filmes

Cada usuário contribuiu com diversas avaliações de diversos filmes. Para cada usuário foi aplicado o algoritmo C4.5 afim de gerar árvores de decisão que classificava entre filmes que é ou não de seu gosto. Assim como na base de dados da sala IoT, esta base de dados foi dividida em conjunto de treinamento e conjunto de teste. Os atributos

de entrada utilizados para aplicar o algoritmo C4.5 foram: Ano de Lançamento, Gênero 1, Gênero 2, Duração, Diretor Ganhou Oscar, Ator Principal Ganhou Oscar, País de Origem, Filme Ganhou Oscar e Nota IMDB. Como atributo de saída se tinha a avaliação do usuário.

Nesta base de dados, a grande maioria dos usuários não apresentavam avaliações balanceadas, ou seja, a quantidade de filmes que agradava geralmente era muito superior a quantidade de filmes que não agradava. Portanto o algoritmo C4.5 foi aplicado em duas situações diferentes:

- a) dados balanceados: a quantidade de filmes que agradaram e que não agradaram são iguais no conjunto de treinamento;
- b) dados desbalanceados: a quantidade de filmes que agradaram é diferente da quantidade de filmes que não agradaram no conjunto de treinamento.

A quantidade de instâncias presentes no conjunto de teste é aproximadamente 20% da quantidade de instâncias no conjunto de treinamento para cada usuário.

Para cada usuário, descritos na tabela 2 e 3, foi gerado uma árvore de decisão. Nesta árvore eram determinadas quais características de filmes que mais agradavam o usuário em questão. Posteriormente a determinação destas árvores, foi testado com o conjunto de testes as porcentagens de acerto, ou seja, quantidades de classificações (agradar ou não agradar) corretas de filmes. Note que na tabela 2 os dados são balanceados e na tabela 3 não. Todas as árvores de decisão geradas estão disponíveis e podem ser visualizadas no repositório github.com/Nahmir/SPRSC.

Em suma, as árvores de decisão geradas para cada usuário atuam como um modelo de preferências para filmes, ou seja, cada filme presente no sistema e que possa ser sugerido, é comparado com as características destas árvores e caso seja classificado como um filme que agrada o usuário o sistema faz esta recomendação.

7.3 VALOR DE LIMIAR

Após determinado a árvore de decisão que determina os momentos de pertinência para se assistir filmes em uma sala IoT hipotética e as árvores de decisão que representam as preferências individuais dos usuários deste SPRSC é necessário determinar de que forma e quando o sistema faz uma recomendação de forma proativa (autônoma).

Tabela 2 - Preferências de Filmes para Instâncias Balanceadas

ID de Usuário	Filmes que agradaram		Acertos com Árvore Podada	Acertos com Árvore Não Podada
	Sim	Não	Conjunto de Teste	Conjunto de Teste
2	30	30	66%	66%
3	10	10	37%	50%
5	70	70	54%	42%
6	15	15	50%	33%
7	-	-	-	-
11	50	50	56%	61%
16	10	10	83%	83%
17	30	30	50%	46%
19	100	100	47%	47%
22	100	100	71%	74%
24	20	20	75%	67%
31	40	40	55%	55%
44	60	60	54%	62%
51	10	10	75%	75%
62	210	210	52%	60%
88	8	8	70%	70%
95	30	30	60%	53%
101	5	5	17%	17%
111	20	20	64%	64%
137	70	70	75%	71%
138	-	-	-	-
280	10	10	60%	60%
314	100	100	55%	32%

Fonte: Autor

Tabela 3 - Preferências de Filmes para Instâncias Desbalanceadas

ID de Usuário	Filmes que agradaram		Acertos com Árvore Podada	Acertos com Árvore Não Podada
	Sim	Não	Conjunto de Teste	Conjunto de Teste
2	50	30	63%	59%
3	29	10	50%	50%
5	70	104	46%	38%
6	44	15	50%	33%
7	26	5	-	-
11	-	-	-	-
16	10	19	67%	67%
17	153	30	54%	64%
19	128	100	50%	46%
22	100	162	71%	76%
24	104	20	58%	75%
31	54	40	45%	55%
44	107	60	58%	73%
51	22	10	50%	0
62	237	210	52%	56%
88	50	8	50%	60%
95	39	30	50%	47%
101	95	5	50%	50%
111	42	21	55%	59%
137	103	70	79%	79%
138	40	4	90%	95%
280	30	10	50%	50%
314	170	100	50%	41%

Fonte: Autor

Para este SPRSC, o sistema seguirá algumas instruções pré-determinadas:

- a) o sistema espera que o contexto da sala IoT seja condizente com um momento pertinente para se assistir filme;
- b) assim que o momento da sala for pertinente, o sistema verifica se um filme em específico possui classificação condizente com o usuário do sistema, ou seja, ele verifica se o usuário tem idade suficiente para assistir ao filme em específico. Caso não tenha ele verifica outro filme;
- c) após verificado a classificação, o sistema encontra um filme que seja de preferência do usuário em específico (conforme a árvore de decisão gerada); e
- d) por fim o sistema faz a sugestão deste filme em específico de forma proativa.

7.4 TÉCNICA DE SISTEMA DE RECOMENDAÇÃO

Para este SPRSC, foi escolhido a técnica Baseado em Conteúdo de Sistema de Recomendação. Foi feita esta escolha devido os usuários do sistema apresentar diversas avaliações de diversos filmes e estes filmes apresentarem diferentes características. Uma vez gerado os modelos de preferências de filmes dos usuários (através das árvores de decisão) é possível prever, para filmes que o usuário não tenha assistido ainda, se estes filmes seriam do agrado ou não.

7.5 AVALIAÇÃO DO SPRSC E CONSIDERAÇÕES FINAIS DA SEÇÃO

Dado o SPRSC de sugestões de filmes acima, este foi avaliado conforme o modelo de avaliação de SPRSC desenvolvido nesta dissertação. Os quadros abaixo detalham cada aspecto de acordo com cada estrutura do sistema.

Quadro 12 – Estrutura Usuários do Modelo de Avaliação

Usuários	
Conhecido	Os usuários conhecidos apresentam diversas avaliações de diversos filmes. Como dados de cunho pessoal, os usuários apresentam: profissão, faixa etária, cidade e estado de onde moram. Como este sistema não foi aplicado na prática, estes usuários não apresentam dados do contexto onde estão.
Não Conhecido	-

Fonte: Autor

Quadro 13 - Estrutura Contexto em Internet das Coisas do Modelo de Avaliação

Contexto em Internet das Coisas	
Aquisição de Dados	<p>O contexto deste SPRSC é caracterizado por uma sala IoT hipotética. Os dados que são adquiridos deste contexto através de sensores são: Período da Semana, Companhia, Estado do Sofá, Período do dia, Luminosidade, Temperatura e Estado da Televisão. Em relação aos usuários presentes nesta sala os seguintes dados são obtidos: faixa etária e profissões.</p>
Modelagem de Contexto	<p>Quem interage: usuários conhecidos pelo SPRSC. Estes usuários apresentam dados com os seguintes conteúdos: filmes assistidos e avaliados, dados de cunho pessoal (faixa etária, profissão, cidade e estado) e modelos de preferências.</p> <p>Com o que interage: a interação entre usuário e SPRSC é feita de duas formas. A primeira é de uma forma indireta com a sala IoT, ou seja, através das características que o contexto tem de acordo com o usuário inserida nela. A segunda é de forma direta, ou seja, através de alguma interface (como exemplo dispositivos móveis), onde as recomendações seriam apresentadas de forma proativa para o usuário. Note que como o sistema não foi aplicado de forma prática estas interações não são originadas ainda.</p> <p>Onde interage: a interação entre o usuário e o SPRSC é feita dentro de uma sala IoT hipotética. Lembrando que nesta existem diferentes sensores.</p> <p>Quando interage: as métricas para aferição de quando existe a interação são período da semana e período do dia.</p> <p>Porque interage: a primeira razão para interação é entretenimento, ou seja, o usuário busca recomendações de filmes que o agrada. A segunda é por motivos profissionais, ou seja, o usuário assiste filmes com intuito de elaborar críticas especializadas.</p>

Fonte: Autor

Quadro 14 - Estrutura Aprendizado de Máquina do Modelo de Avaliação

Aprendizado de Máquina	
Métodos	<p>Modelos de Preferência: é utilizado o algoritmo C4.5 nas diferentes avaliações de diferentes filmes para gerar árvores de decisão para cada usuário do SPRSC. Estas árvores compreendem as características de filmes que mais agradam os usuários.</p> <p>Momentos de Pertinência: é utilizado o algoritmo C4.5 na base de dados da sala IoT hipotética. Este algoritmo gera uma árvore de decisão que determina quais momentos são propícios para assistir um filme. Esta árvore compreende as características do contexto em que os usuários assistiriam a um filme qualquer.</p>
Regras	Para este SPRSC que faz sugestões de filmes a regra que ele compreende é em relação a classificação indicativa para cada filme.
Valor de Limiar	O valor de limiar é determinado através da relação de momento de pertinência para se assistir um filme na sala IoT hipotética, classificação indicativa de um filme em específico e o modelo de preferência de um usuário qualquer.

Fonte: Autor

Quadro 15 - Estrutura Sistemas de Recomendação do Modelo de Avaliação

Sistemas de Recomendação	
Individual	Neste SPRSC é utilizado de uma abordagem denominada Baseado em Conteúdo.
Híbrido	-

Fonte: Autor

Ao longo desta dissertação, diversos conceitos e pesquisas a respeito das áreas de Sistemas de Recomendação, Sistemas de Recomendação Sensíveis ao Contexto, Internet das Coisas e Aprendizado de Máquina foram debatidos. Destas pesquisas, notou-se que sistemas que buscam autonomia e aumento da qualidade para fornecer suas sugestões necessitam de diversas relações entre estas áreas. Afim de verificar como são feitas as relações destas diversas áreas em Sistemas Proativos de Recomendação, foi elaborado nesta dissertação um modelo de avaliação para Sistemas Proativos de Recomendação

Sensíveis ao Contexto. Como dito anteriormente, este modelo é separado em quatro estruturas, sendo elas: Contexto em Internet das Coisas, Aprendizado de Máquinas, Usuários e Sistemas de Recomendação.

A primeira conclusão que se chega ao estudar SPRSC, é que são necessários dados do contexto para determinar modelos de preferências e também determinar quais os momentos propícios para uma sugestão autônoma de acordo com as necessidades momentâneas dos usuários compreendidos pelo sistema proativo. Para isso, conforme visto na seção 3, a IoT através de seus inúmeros dispositivos que conseguem modelar objetos do mundo físico em objetos do mundo virtual, permite que estes sistemas tenham acesso aos dados do contexto onde estão inseridos usuários e sistemas. A IoT se torna uma grande aliada e parte integrante de SPRSC devido suas primícias compreenderem as seguintes finalidades: comunicação de qualquer coisa, em qualquer lugar e a qualquer momento. Sendo assim, grandes quantidades de dados e de diversos tipos ficam disponíveis para que os SPRSC usem e elaborem suas aprendizagens para posteriormente sugerir itens, ações e/ou serviços de forma proativa e que atendam às necessidades de seus usuários.

Em um ambiente real e que compreenda as finalidades da IoT, diversos dados podem ser compreendidos pelos SPRSC, porém existem situações que nem todos são necessários para caracterizar o contexto onde estão inseridos diretamente os usuários e/ou sistema. Uma forma para identificar quais dados auxiliam de forma direta para caracterização do contexto é utilizar daqueles dados que possam responder uma ou mais das seguintes questões: quem interage, com o que interage, onde é feita a interação, quando é feita a interação e porque é feita a interação. Lembrando que estas perguntas para caracterização do contexto foram analisadas e explicadas na seção 3 desta dissertação.

Outra questão que vale a pena ser exposta quando se trata de SPRSC que utilizam da IoT para obter dados do contexto onde usuários e sistema estão inseridos, está relacionado as métricas para obter preferências e necessidades de seus usuários. No Sistema Proativo de Recomendação de filmes simulado nesta dissertação, uma sala IoT (ou seja, uma sala que tem diversos sensores que tenham a capacidade de comunicar seu estado) hipotética para se assistir filmes foi simulada. Conforme descrito na subseção 7.1.1, esta sala teria diversos sensores que conseguiram aferir diversas características do contexto. Perguntas aleatórias que retratavam diferentes situações de contexto, ou seja, diferentes valores para os sensores teoricamente simulados, foram feitas a diversas

peessoas afim de determinar quais eram as melhores características do contexto que determinavam os momentos propícios para se assistir um filme. Note que devido à sala IoT ter sido simulada teoricamente, as preferências para determinar quais momentos eram pertinentes para se assistir filmes foram obtidos de forma explícita. Ao implementar uma sala IoT na prática, ou seja, com sensores reais e que mandem realmente dados do contexto, os SPRSC podem monitorar as ações dos usuários dentro desta sala IoT e apreender através de métodos de aprendizagem de máquina quais são os momentos em que estes usuários assistem filmes. Note que nestas situações, modelos de preferências são determinados de forma implícita, sendo esta uma característica proveniente da IoT que pode ser integrada aos Sistemas Proativo de Recomendação. Conforme debatido na subseção 5.1, preferências explícitas exigem que os usuários depreendam esforços para mandar suas avaliações, sendo que em muitas situações os usuários não as fornecem. Já avaliações implícitas, não necessitam de esforços da parte dos usuários pois as ações dos mesmos são monitoradas. Nestes casos, modelos de preferências se tornam muito subjetivos e necessitam de métricas mais apuradas para avaliações.

O segundo ponto a se destacar em relação aos SPRSC envolve as tratativas para obtenção de conhecimento a partir dos diferentes dados contemplados pelo sistema. Conforme explicado na seção 2 desta dissertação, Sistemas Proativos de Recomendação por definição devem fornecer sugestões de forma autônoma, ou seja, sem uma solicitação explícita do usuário. Para isso, três aspectos devem ser desenvolvidos e relacionados entre si pelos sistemas, sendo eles: modelos de preferências, identificação de momentos de pertinência e por fim as regras que descrevem o contexto de acordo com a finalidade do sistema proativo.

Modelos de preferências, como o próprio nome diz, está relacionado a identificar quais são os gostos ou necessidades dos usuários inseridos nos sistemas proativos. Estes modelos são obtidos através das avaliações explícitas e/ou implícitas dos usuários dadas aos itens, ações e/ou serviços já utilizados. No SPRSC de filmes desenvolvido nesta dissertação e explicado nesta seção, para gerar os modelos de preferências dos usuários foi utilizado de um método supervisionado de aprendizado de máquina denominado algoritmo C4.5, onde de acordo com as diferentes avaliações para os diferentes filmes já assistidos por um usuário em específico, eram obtidas árvores de decisão que levantavam as características de filmes que mais impactavam nos gostos deste usuário em específico.

Referente aos momentos de pertinência, estes estão relacionados as situações (características do contexto) que apresentam um grande potencial para uma sugestão

autônoma pelo sistema proativo. Igual a forma utilizada para obtenção dos modelos de preferências, também foi utilizado o algoritmo C4.5 nos dados referentes a sala IoT hipotética afim de identificar quais características do contexto descreviam um momento de pertinência para uma sugestão autônoma do SPRSC de filmes. Ao aplicar o algoritmo nestes dados, árvores de decisão eram geradas identificando quais características de contexto da sala IoT hipotética tinham mais impacto para definir um momento propício para se assistir um filme.

Por fim, os diferentes SPRSC com diferentes finalidades devem compreender quais aspectos que descrevem as regras do contexto onde estão inseridos, ou seja, quais são as diretrizes que devem ser respeitadas para que as sugestões autônomas não sejam abusivas, errôneas ou até mesmo desrespeitosas. O objetivo do sistema simulado nesta dissertação é fazer sugestões autônomas de filmes. Quando se trata de filmes, a regra que descreve o contexto para sugerir um filme qualquer é primeiro verificar se o usuário que irá receber a sugestão autônoma tem idade suficiente para se assistir ao filme de acordo com a classificação indicada.

Uma vez determinado estes três aspectos pelo sistema, eles devem ser relacionados afim de gerar valores de limiar que devem ser superados para que o sistema proativo faça uma sugestão de forma autônoma. No caso do SPRSC de filmes simulado, o valor de limiar é explicado mais detalhadamente na subseção 7.3.

Ao relacionar as estruturas Contexto em IoT e Aprendizado de Máquina do modelo de avaliação de SPRSC e após diversas análises de pesquisas compreendidas nesta dissertação, outros pontos relevantes merecem destaque para reflexão. Um Sistema Proativo de Recomendação que está inserido em um ambiente IoT, pode aprender momentos de pertinência de um ambiente coletivo e/ou de um ambiente que tenham sempre os mesmos indivíduos. Como exemplo desta situação, a sala IoT para se assistir filmes simulada nesta dissertação, compreendia diversos usuários que não tinham nenhuma relação entre si e conseqüentemente tinham suas preferências de contexto muitas vezes diferentes umas das outras. Note que ao aplicar um método de aprendizado de máquina para determinar os momentos de pertinência nesta sala, a aprendizagem se dá a partir de dados de um ambiente coletivo, ou seja, as preferências de contexto são determinadas através de um grupo de pessoas e não de uma em específico. Momentos de pertinência obtidos a partir de aprendizado de ambientes coletivos se mostra útil, quando o SPRSC utiliza deste aprendizado para fornecer uma sugestão para usuários que apresentam pouco ou até mesmo nenhum dado sobre suas preferências de contexto, ou

seja, o sistema proativo mesmo que não saiba as preferências deste novo usuário, ele não fica incapacitado para determinar um momento propício para uma sugestão autônoma. Quando o aprendizado é feito a partir dos dados de um grupo específico de usuários em ambientes IoT, os momentos de pertinência aprendidos são mais personalizados, portanto, os SPRSC conseguem ter maior qualidade para indicar quais momentos são mais propícios para uma sugestão para estes usuários.

Da situação apresentada no parágrafo acima, depreende-se outra característica que deve ser avaliada nos diversos Sistemas Proativos de Recomendação. Conforme demonstrado na subestrutura Usuários do modelo de avaliação de SPRSC, existem dois tipos de usuários que podem utilizar os sistemas, sendo eles: conhecidos e não conhecidos. Usuários conhecidos apresentam diversos dados em relação ao contexto onde estão inseridos, dados de cunho pessoal e diversas avaliações dos itens, ações e/ou serviços já utilizados. Note que para estes usuários é possível que os SPRSC apliquem métodos de aprendizado de máquina para aprender quais são as preferências em relação a momentos de pertinência para uma sugestão proativa e quais itens, ações e/ou serviços recomendar. Porém, existem situações em que os SPRSC terão de fornecer sugestões a usuários não conhecidos, ou seja, usuários que não apresentam dados do contexto onde estão inseridos, dados de cunho pessoal e avaliações de itens, ações e/ou serviços já utilizados. Portanto, ao avaliar um SPRSC é necessário que sejam determinados quais métricas são utilizadas para que o sistema forneça sugestões para usuários não conhecidos e quais são os dados que caracterizam um usuário conhecido.

Como Sistemas Proativos de Recomendação surgiram a partir das abordagens mais tradicionais de Sistemas de Recomendação (explicados na seção 4), a fase para identificar um item, ação e/ou serviço da grande maioria destes sistemas se baseiam nas técnicas individuais e/ou híbridas dos Sistemas de Recomendação. Como exemplo prático, o SPRSC de filmes desenvolvido nesta dissertação, baseava suas recomendações na técnica individual de SR Baseados em Conteúdo. Esta técnica individual foi escolhida devido o SPRSC aplicar seu aprendizado de máquina nas diferentes avaliações de filmes provenientes de diferentes usuários para determinar quais as características de filmes mais os agradavam. Após determinadas estas características de preferência, o sistema verificava quais filmes apresentavam características semelhantes e posteriormente fazia suas sugestões.

Em suma, após conceitos explicados, pesquisas analisadas e simulações feitas nesta dissertação, conclui-se que Sistemas Proativos de Recomendação Sensíveis ao

Contexto devem compreender características bem definidas para terem eficácia e serem prestativos para seus usuários. Estas características são identificadas no modelo de avaliação de SPRSC, permitindo que interessados nas tratativas e relações existentes nos mais diversos sistemas proativos, possam fazer uma análise mais aprofundada e conseqüentemente ter um aprendizado mais eficaz e organizado.

8 CONCLUSÃO E TRABALHOS FUTUROS

Após desenvolvida esta dissertação, de acordo com os estudos, desenvolvimentos e simulações de Sistemas Proativos de Recomendação Sensíveis ao Contexto, alguns pontos devem ser destacados e evidenciados. Vale ressaltar também alguns itens que podem ser desenvolvidos no futuro e que auxiliam no desenvolvimento do conceito de Sistemas Proativos de Recomendação.

8.1 CONCLUSÃO

A IoT permite que diversos dados e de diferentes tipos sejam originados. Quando analisados de forma individual e/ou conjuntas, permitem que sistemas desenvolvidos para predições sejam capazes de levar em consideração informações do contexto. Como a IoT objetiva criar um mundo melhor para a humanidade, onde as coisas saibam o que os seres humanos gostam, desejam, precisam e hajam de forma coerente sem instruções explícitas, questões como autonomia de sistemas inseridos na IoT cada vez mais são discutidas. Outro objetivo da IoT é determinar o que os seres humanos gostam, desejam e precisam. Estas questões podem ser inseridas em diversos grupos de aplicações, tais como: *Smart city*, *Smart Water*, *Smart Farm*, entre outros.

Ao unir a IoT e Sistemas de Recomendação, a funcionalidade mais importante que se origina é a sensibilidade ao contexto. Conforme apresentado nesta dissertação, contexto é definido como qualquer dado e/ou informação que possa ser usada para caracterizar a situação de uma entidade. Através dos sensores que compreendem as primícias da IoT, estes dados e/ou informações são gerados e obtidos. Estes dados e/ou informações quando utilizados por Sistemas de Recomendação, permitem que suas sugestões tenham mais eficácia e sejam mais personalizadas para seus usuários. Ao conhecer o contexto onde o usuário e Sistema de Recomendação estão inseridos, através de métodos de aprendizado de máquina aplicados a estes dados originados da IoT, é possível determinar momentos em que o usuário necessita ou aprecia uma sugestão, sendo assim, o conceito de Sistemas Proativos de Recomendação Sensível ao Contexto surge.

Nesta dissertação, através da leitura e análise de diferentes artigos e pesquisas, foi levantado quais estruturas e características mais impactantes devem ser analisadas e avaliadas quando se trata de Sistemas Proativos de Recomendação Sensíveis ao Contexto.

Ao se aplicar o modelo de avaliação para Sistemas Proativos de Recomendação Sensíveis ao Contexto desenvolvido neste trabalho, as questões mais preocupantes e que são de fundamental importância para o desenvolvimento destes sistemas, são diagnosticados e mostram quais são as tratativas que o sistema faz para elaborar e sugerir recomendações de forma autônoma. Lembrando que estas recomendações proativas devem atender realmente as necessidades de seus usuários nos momentos certos.

Outra conclusão importante a se destacar, é em relação ao método de aprendizado de máquina utilizado nesta dissertação para a simulação de um SPRSC de filmes. Devido à natureza das bases de dados utilizadas neste sistema, onde os dados eram classificados, o algoritmo de árvore de decisão C4.5 foi utilizado e conforme os resultados obtidos e descritos na simulação, apresentou bons índices tanto para a determinação de momentos propícios para se assistir filme em uma sala IoT, como para traçar um perfil individual de usuário em relação a preferências de filmes.

Em suma, o modelo de avaliação desenvolvido utiliza de informações provenientes dos sensores inseridos na IoT e quando tratados de forma individual e/ou conjunta, permitem a caracterização do contexto onde os SPRSC estão inseridos. Portanto este modelo atingiu seu objetivo principal que era determinar quais são as principais características de Sistemas Proativos de Recomendação inseridos no âmbito de IoT.

8.2 TRABALHOS FUTUROS

Um ponto importante como trabalho futuro, é aplicar o SPRSC simulado nesta dissertação a uma sala IoT real e desenvolver como interface entre usuário e sistema, uma aplicação mobile que forneça sugestões de filmes de forma proativa. Esta aplicação seria importante para abordar outros temas que não foram discutidos nesta dissertação, tais como: de que forma a interface *mobile* deve apresentar as recomendações proativas, como desenvolver ferramentas que permitam *feedbacks*, de que forma estes *feedbacks* influenciam nos valores de limiar do sistema e em seu aprendizado e como traçar perfis comportamentais de usuários ao longo das suas interações com o sistema.

Outro ponto importante é considerar outros métodos de aprendizado de máquina ao se desenvolver Sistemas Proativos de Recomendação, tais como: modelos não supervisionados, semi supervisionados e por reforço. Através destes métodos, determinar quais algoritmos apresentam melhores resultados de acordo com diversos contextos e quantidade de informações.

Em relação a IoT, determinar métricas que tratam da subjetividade de avaliações indiretas de itens, ações e/ou serviços de acordo com o comportamento do usuário em um ambiente com diversos sensores.

Por fim, utilizando deste modelo de avaliação de SPRSC e das estruturas fundamentais levantadas nesta dissertação, desenvolver Sistemas Proativos de Recomendação que auxiliem seus usuários no sentido de preservação e economia de recursos naturais, sendo que o usuário deve ter a menor intervenção possível.

REFERÊNCIAS

ABOWD, Gregory D.; MYNATT, Elizabeth D. Charting past, present, and future research in ubiquitous computing. **ACM Transactions on Computer-Human Interaction (TOCHI)**, v. 7, n. 1, p. 29-58, 2000.

ADOMAVICIUS, Gediminas; TUZHILIN, Alexander. 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**, v. 17, n. 6, p. 734-749, 2005.

ALCARAZ, Cristina et al. Wireless sensor networks and the internet of things: Do we need a complete integration?. In: **1ST INTERNATIONAL WORKSHOP ON THE SECURITY OF THE INTERNET OF THINGS (SECIOT'10)**. 2010.

AL-FUQAHA, Ala et al. Internet of things: A survey on enabling technologies, protocols, and applications. **IEEE Communications Surveys & Tutorials**, v. 17, n. 4, p. 2347-2376, 2015.

ASGHAR, Mohsen Hallaj; NEGI, Atul; MOHAMMADZADEH, Nasibeh. Principle application and vision in Internet of Things (IoT). In: **Computing, Communication & Automation (ICCCA), 2015 International Conference on**. IEEE, 2015. p. 427-431.

ASHTON, Kevin. That 'internet of things' thing. **RFiD Journal**, v. 22, n. 7, p. 97-114, 2009.

BARR, Avron; FEIGENBAUM, E. A. **The Handbook of Artificial Intelligence**, vol. 1. Los Altos, CA: VWilliam Kaufman, 1981.

BASGALUPP, Márcio Porto; DE CARVALHO, André CPLF; FREITAS, Alex A. **LEGAL-Tree: Um algoritmo genético multi-objetivo lexicográfico para indução de árvores de decisão**. 2010. Tese de Doutorado. Tese de Doutorado, ICMC-USP, São Carlos.

BAZIRE, Mary; BRÉZILLON, Patrick. Understanding context before using it. In: **INTERNATIONAL AND INTERDISCIPLINARY CONFERENCE ON MODELING AND USING CONTEXT**. Springer Berlin Heidelberg, 2005. p. 29-40.

BEDI, Punam; AGARWAL, Sumit Kr. A situation-aware proactive recommender system. In: **HYBRID INTELLIGENT SYSTEMS (HIS), 2012 12TH INTERNATIONAL CONFERENCE ON**. IEEE, 2012. p. 85-89.

BEDI, Punam et al. Sapr: Situation-aware proactive recommender system with explanations. In: **Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on**. IEEE, 2014. p. 277-283.

BIELSA, Alberto. **Measuring the inertial forces in two of the most savage roller coasters of the world with wireless sensors**. 2012a. Disponível em: <http://www.libelium.com/wireless_sensors_measure_inertial_forces_accelerometer/>. Acesso em: 20 dez. 2016.

BIELSA, Alberto. **Smart Cars: a practical implementation of M2M communications is becoming a reality ever closer.** 2012b. Disponível em: <http://www.libelium.com/smart_cars_m2m_accident_prevention>. Acesso em: 20 dez. 2016.

BOBADILLA, Jesús et al. **Recommender systems survey.** *Knowledge-Based Systems*, v. 46, p. 109-132, 2013.

BRAUNHOFER, Matthias et al. A context-aware model for proactive recommender systems in the tourism domain. In: **PROCEEDINGS OF THE 17TH INTERNATIONAL CONFERENCE ON HUMAN-COMPUTER INTERACTION WITH MOBILE DEVICES AND SERVICES ADJUNCT.** ACM, 2015. p. 1070-1075.

BURANGE, Anup W.; MISALKAR, Harshal D. Review of Internet of Things in development of smart cities with data management & privacy. In: **Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in.** IEEE, 2015. p. 189-195.

BURKE, Robin. Hybrid recommender systems: Survey and experiments. **User modeling and user-adapted interaction**, v. 12, n. 4, p. 331-370, 2002.

DEY, Anind K.; ABOWD, Gregory D.; SALBER, Daniel. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. **Human-computer interaction**, v. 16, n. 2, p. 97-166, 2001.

EVANS, Dave. The internet of things. **How the Next Evolution of the Internet is Changing Everything, Whitepaper, Cisco Internet Business Solutions Group (IBSG)**, v. 1, p. 1-12, 2011.

FLEISCH, Elgar et al. What is the internet of things? An economic perspective. **Economics, Management, and Financial Markets**, n. 2, p. 125-157, 2010.

GALLEGO, Daniel et al. Incorporating proactivity to context-aware recommender systems for e-learning. In: **Computer and Information Technology (WCCIT), 2013 World Congress on.** IEEE, 2013. p. 1-6.

GALLEGO, Daniel; WOERNDL, Wolfgang; HUECAS, Gabriel. Evaluating the impact of proactivity in the user experience of a context-aware restaurant recommender for Android smartphones. **Journal of Systems Architecture**, v. 59, n. 9, p. 748-758, 2013.

GASCÓN, David; YARZA, Marcos. **Wireless Sensor Networks to Control Radiation Levels.** 2011. Disponível em: <http://www.libelium.com/wireless_sensor_networks_to_control_radiation_levels_geiger_counters/>. Acesso em: 20 dez. 2016.

GUBBI, Jayavardhana et al. Internet of Things (IoT): A vision, architectural elements, and future directions. **Future Generation Computer Systems**, v. 29, n. 7, p. 1645-1660, 2013.

INTERNATIONAL TELECOMMUNICATION UNION. **Y.2060**: Overview of the Internet of things. Geneva, 2012. 14 p.

IOTSENS GRUPO GIMENO IOT DIVISION (Org.). **CASTELLÓN SMART CITY – PILOT**. Disponível em: <<http://www.iotsens.com/castellon-smart-city-pilot/>>. Acesso em: 12 dez. 2016.

LIBELIUM. Cold Chain Monitoring and Smart Tracking in Zaragoza. 2015a. Disponível em: <<http://www.libelium.com/cold-chain-monitoring-smart-tracking-zaragoza/>>. Acesso em: 20 dez. 2016.

LIBELIUM. Fish farm monitoring in Vietnam by controlling water quality in ponds and tanks. 2016a. Disponível em: <<http://www.libelium.com/fish-farm-monitoring-in-vietnam-by-controlling-water-quality-in-ponds-and-tanks/>>. Acesso em: 20 dez. 2016.

LIBELIUM. Precision Agriculture: Predicting Vineyard Conditions, Preventing Disease. 2015b. Disponível em: <<http://www.libelium.com/precision-agriculture-predicting-vineyard-conditions-preventing-disease/>>. Acesso em: 20 dez. 2016.

LIBELIUM. Smart Farming: Monitoring Horses and Equine Facility Management with Waspnote. 2015c. Disponível em: <<http://www.libelium.com/smart-farming-monitoring-horses-equine-facility-management-waspnote/>>. Acesso em: 20 dez. 2016.

LIKA, Blerina; KOLOMVATSOS, Kostas; HADJIEFTHYMIADES, Stathes. Facing the cold start problem in recommender systems. **Expert Systems with Applications**, v. 41, n. 4, p. 2065-2073, 2014.

LOPS, Pasquale; DE GEMMIS, Marco; SEMERARO, Giovanni. Content-based recommender systems: State of the art and trends. **Recommender systems handbook**. Springer US, 2011. p. 73-105.

MANYIKA, James et al. **Disruptive technologies**: Advances that will transform life, business, and the global economy. San Francisco, CA: McKinsey Global Institute, 2013.

MASHAL, Ibrahim; ALSARYRAH, Osama; CHUNG, Tein-Yaw. Performance evaluation of recommendation algorithms on Internet of Things services. **Physica A: Statistical Mechanics and its Applications**, v. 451, p. 646-656, 2016.

MCEWEN, Adrian; CASSIMALLY, Hakim. **Designing the internet of things**. John Wiley & Sons, 2013.

MICHAELIS. **Dicionário Brasileiro da Língua Portuguesa**. 2015. Disponível em: <<http://michaelis.uol.com.br/busca?id=EPPG>>. Acesso em: 09 mar. 2017.

MITCHELL, Tom M. **Machine learning**. 1997. Burr Ridge, IL: McGraw Hill, v. 45, 1997.

MONARD, Maria Carolina; PRATI, Ronaldo Cristiano. **Aprendizado de Máquina Simbólico para Mineração de Dados**, 200?.

MUNOZ-ORGANERO, Mario et al. A collaborative recommender system based on space-time similarities. **IEEE Pervasive Computing**, v. 9, n. 3, p. 81-87, 2010.

PAZZANI, Michael J. A framework for collaborative, content-based and demographic filtering. **Artificial Intelligence Review**, v. 13, n. 5-6, p. 393-408, 1999.

PAZZANI, Michael J.; BILLSUS, Daniel. Content-based recommendation systems. **The adaptive web**. Springer Berlin Heidelberg, 2007. p. 325-341.

PERERA, Charith et al. Context aware computing for the internet of things: A survey. **IEEE Communications Surveys & Tutorials**, v. 16, n. 1, p. 414-454, 2014.

PONISZEWSKA-MARANDA, Aneta; KACZMAREK, Daniel. Selected methods of artificial intelligence for Internet of Things conception. In: **Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on**. IEEE, 2015. p. 1343-1348.

QUINLAN, J. **C4. 5: Programs for Machine Learning**. C4. 5-programs for machine learning. 1993 OK.

RIGGINS, Frederick J.; WAMBA, Samuel Fosso. Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics. In: **System Sciences (HICSS), 2015 48th Hawaii International Conference on**. IEEE, 2015. p. 1531-1540.

ROKACH, Lior; MAIMON, Oded. **Data mining with decision trees: theory and applications**. World scientific, 2014.

RUSSELL, Stuart; NORVIG, Petter. **Inteligência Artificial**, 3ª Edição. Elsevier Brasil, 2014.

SABIC, Adem. Proactive Recommendation Delivery. In: **PROCEEDINGS OF THE 10TH ACM CONFERENCE ON RECOMMENDER SYSTEMS**. ACM, 2016. p. 459-462.

SACHDEVA, Nitin; DHIR, Renu; KUMAR, Akshi. Empirical analysis of Machine Learning Techniques for context aware Recommender Systems in the environment of IoT. In: **PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ADVANCES IN INFORMATION COMMUNICATION TECHNOLOGY & COMPUTING**. ACM, 2016. p. 39.

SALMAN, Yassmeen et al. A Proactive Multi-type Context-Aware Recommender System in the Environment of Internet of Things. In: **COMPUTER AND INFORMATION TECHNOLOGY; UBIQUITOUS COMPUTING AND COMMUNICATIONS; DEPENDABLE, AUTONOMIC AND SECURE COMPUTING; PERVASIVE INTELLIGENCE AND COMPUTING (CIT/IUCC/DASC/PICOM), 2015 IEEE INTERNATIONAL CONFERENCE ON**. IEEE, 2015. p. 351-355.

SANCHEZ, Luis et al. A generic context management framework for personal networking environments. In: **MOBILE AND UBIQUITOUS SYSTEMS-WORKSHOPS, 2006. 3RD ANNUAL INTERNATIONAL CONFERENCE ON**. IEEE, 2006. p. 1-8.

SCHAFER, J. Ben et al. Collaborative filtering recommender systems. In: **THE ADAPTIVE WEB**. Springer Berlin Heidelberg, 2007. p. 291-324.

SCHILIT, Bill; ADAMS, Norman; WANT, Roy. Context-aware computing applications. In: **MOBILE COMPUTING SYSTEMS AND APPLICATIONS, 1994. WMCSA 1994. FIRST WORKSHOP ON**. IEEE, 1994. p. 85-90.

SGARBI, Julio André. **Domótica Inteligente: Automação residencial baseada em comportamento**. 2007. 107 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Centro Universitário Fei, São Bernardo do Campo, 2007. Disponível em: <<http://fei.edu.br/~flaviot/ibas/index.htm>>. Acesso em: 10 jan. 2017.

SIMON, Herbert A. Why should machines learn?. In: **Machine learning**. Springer Berlin Heidelberg, 1983. p. 25-37.

SINGH, Sachchidanand; SINGH, Nirmala. Internet of Things (IoT): Security challenges, business opportunities & reference architecture for E-commerce. In: **GREEN COMPUTING AND INTERNET OF THINGS (ICGCIOT), 2015 INTERNATIONAL CONFERENCE ON**. IEEE, 2015. p. 1577-1581.

SOLOBERA, Javier. **Detecting Forest Fires using Wireless Sensor Networks**. 2010. Disponível em: <http://www.libelium.com/wireless_sensor_networks_to_detec_forest_fires>. Acesso em: 12 dez. 2016.

TSAI, Chun-Wei; LAI, Chin-Feng; VASILAKOS, Athanasios V. Future Internet of Things: open issues and challenges. **Wireless Networks**, v. 20, n. 8, p. 2201-2217, 2014.

VON WANGENHEIM, Christiane Gresse; VON WANGENHEIM, Aldo. **Raciocínio baseado em casos**. Editora Manole Ltda, 2003.

WEISER, Mark. The computer for the 21st century. **Scientific american**, v. 265, n. 3, p. 94-104, 1991.

WITTEN, Ian H. et al. **Data Mining: Practical machine learning tools and techniques**. Morgan Kaufmann, 2016.

WOERNDL, Wolfgang et al. A model for proactivity in mobile, context-aware recommender systems. In: **PROCEEDINGS OF THE FIFTH ACM CONFERENCE ON RECOMMENDER SYSTEMS**. ACM, 2011. p. 273-276.

YAO, Lina et al. Exploring recommendations in internet of things. In: **PROCEEDINGS OF THE 37TH INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH & DEVELOPMENT IN INFORMATION RETRIEVAL**. ACM, 2014. p. 855-858.