



UNIVERSIDADE FEDERAL DA BAHIA

TRABALHO DE GRADUAÇÃO

**Deteccção inteligente de efeitos colaterais indesejáveis na
Internet das coisas - um estudo de caso no Home Network
System**

Heron Sanches Gonçalves Pires Ferreira

Programa de Graduação em Ciência da Computação

Salvador
31 de outubro de 2016

HERON SANCHES GONÇALVES PIRES FERREIRA

**DETECÇÃO INTELIGENTE DE EFEITOS COLATERAIS
INDESEJÁVEIS NA INTERNET DAS COISAS - UM ESTUDO DE
CASO NO HOME NETWORK SYSTEM**

Este Trabalho de Graduação foi apresentado ao Programa de Graduação em Ciência da Computação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Orientadora: Profa. Dra. Daniela Barreiro Claro
Co-orientador: Roberto Cerqueira Figueiredo

Salvador
31 de outubro de 2016

Ficha catalográfica.

Ferreira, Heron Sanches Gonçalves Pires

Detecção inteligente de efeitos colaterais indesejáveis na Internet das coisas - um estudo de caso no Home Network System/ Heron Sanches Gonçalves Pires Ferreira– Salvador, 31 de outubro de 2016.

43p.: il.

Orientadora: Profa. Dra. Daniela Barreiro Claro.

Co-orientador: Roberto Cerqueira Figueiredo.

Monografia (Graduação)– UNIVERSIDADE FEDERAL DA BAHIA, INSTITUTO DE MATEMÁTICA, 31 de outubro de 2016.

“1. Efeitos Colaterais Indesejáveis. 2. Internet das Coisas. 3. Home Network System. 4. Dispositivos como serviços Web RESTful. 5. DECORATE. 6. Cross-validation”.

I. Claro, Daniela Barreiro. II. .

III. UNIVERSIDADE FEDERAL DA BAHIA. INSTITUTO DE MATEMÁTICA. IV. Título.

CDD 005.13307

À minha mãe.

AGRADECIMENTOS

Obrigado minha mãe, por sua dedicação durante todos estes anos. Obrigado minha tia Suzi, por ter me dado oportunidade de estudo e sempre está disposta a me ajudar em qualquer situação. Obrigado as forças boas que vêm me guiando e me ajudando. Obrigado Daniela e Roberto pela orientação prestada para este trabalho.

"Fale para ele não desistir".
—MENSAGEM ESPÍRITA.

RESUMO

Diversos dispositivos têm sido incorporados em ambientes domésticos através dos *Home Server* (HS) com o intuito de automatizar as tarefas cotidianas de uma casa inteligente. Porém, muitos destes dispositivos independentes não conseguem atender aos requisitos dos usuários. Assim, há a necessidade de composição destes com outros dispositivos para prover funcionalidades que atendam a esses requisitos. Quando existe uma composição de dispositivos, o comportamento de pelo menos um destes pode provocar interação de características com efeitos colaterais indesejáveis. Assim, o presente trabalho propõe detectar efeitos colaterais indesejáveis entre dispositivos utilizando-se do *ensemble* DECORATE. Para isto, observou-se o comportamento físico e as mensagens trocadas entre os dispositivos, os quais foram disponibilizados como serviços Web RESTFul. Através do conjunto de dados obtidos realizou-se um processo classificatório e como resultado final foi obtido um modelo com mais de 90% de precisão, o qual foi incorporado no FIM (*Feature Interaction Manager*) do *Home Network System* (HNS), mostrando que é possível detectar efeitos colaterais indesejáveis entre dispositivos.

Palavras-chave: dispositivos, serviços Web RESTFul, *Feature Interaction Manager*, *Home Network System*, efeitos colaterais indesejáveis, *ensemble* DECORATE

ABSTRACT

Many devices have been put on domestic environments through the Home Servers with the aim of automate the daily tasks of a smart home. But, some devices when alone do not attend the user specifications. Then, there is the needed to compose them to provide features that are accord to them specifications. When there is a device's composition, the behave of at least one of them can cause feature interactions with undesirable side effects. So, this work proposes detect undesirable side effects between devices using the ensemble DECORATE. For it propose was realized observations of the physical devices behave and the messages changed by these devices, which were made available how Web RESTFul services. Through the data collected was made a classification process and as the final result was obtained a model with more than 90% of precision. It was put into FIM (Feature Interaction Manager) of the Home Network System (HNS), showing that is possible detect undesirable side effect between these devices.

Keywords: devices, Web RESTFul services, Home Network System, undesirable side effect, Feature Interaction Manager, ensemble DECORATE

SUMÁRIO

Capítulo 1—Introdução	1
Capítulo 2—Fundamentação Teórica	5
2.1 Internet das coisas	5
2.1.1 Dispositivos como serviços Web RESTFul	6
2.2 Interação de características	9
2.3 Home network system	12
2.4 Classificação - Aprendizado Supervisionado	12
2.4.1 DECORATE - um método <i>ensemble</i>	15
2.4.2 Métodos avaliativos	15
Capítulo 3—Proposta	19
3.1 Arquitetura	19
3.1.1 <i>Smartphone Android (1)</i>	21
3.1.2 Raspberry Pi (2), elevador (2.1), garra (2.2) e sensor (2.1.1)	21
3.1.3 PC (3) e Arduino UNO (3.1)	21
3.1.4 Home server - FIM (4)	21
3.1.5 Firebase Cloud Message (6)	21
3.1.6 Amazon RDS	22
3.2 Cenário “Levar compras”	22
3.2.1 Execução do cenário Levar compras	23
3.3 Método de detecção	26
3.3.1 Geração da massa de dados	26
3.3.2 Seleção dos atributos	27
3.3.3 Meta-classificador DECORATE	28
3.4 Implantação do método proposto	29
Capítulo 4—Validação	31
4.1 Experimentos	31
4.2 Resultados e discussão	31
4.2.1 Modelos	31
4.2.2 Implantação	34
Capítulo 5—Trabalhos Relacionados	35

Capítulo 6—Conclusão

LISTA DE FIGURAS

2.1	Dispositivo sensor de obstáculo do elevador disponibilizado como serviço. Faixa de detecção compatível com a largura do elevador utilizado na maquete do cenário (Figura 3.3)	8
13	figure.caption.3	
2.3	(a) Problema linearmente separável. (b) Problema não separável linearmente. (Elizondo, 2006)	14
2.4	k-fold-cross-validation. Adaptado de (Olson e Delen, 2008).	18
3.1	Modelo do HNS utilizado neste trabalho.	20
3.2	Cenário Levar compras.	22
3.3	Dispositivos reais do cenário “Levar compras”.	24
3.4	Tela da aplicação Android do usuário da casa.	24
3.5	Exemplos de cestas do cenário “Levar compras”.	25
3.6	Lista dos objetos que podem ser utilizados para compor uma cesta.	25
3.7	Cesta posicionada no local adequado por um humano, para que a garra possa prender a cesta adequadamente.	26
3.8	Plotagem do espaço de <i>features</i> par a par do dataset de 59 instâncias.	28
3.9	Separação das classes entre os atributos (somaLargura, somaMassa) - “sl x sm”, (somaDiâmetro, somaLargura) - “sd x sl” e, (médiaAltura, somaLargura).	29
3.10	Implantação do método proposto.	30
4.1	<i>Accuracy</i> , <i>Precision</i> e <i>Recall</i> obtidas utilizando stratified-10-fold-crossvalidation (repetido 10 vezes) e DECORATE sobre diferentes frações do <i>dataset</i> com os pares de parâmetros selecionados na seção 3.3.2.	32
4.2	<i>stratified-10-fold-cross-validation</i> (repetido dez vezes). 45 instâncias foram escolhidas randomicamente em cima do <i>dataset</i> de 59 instâncias. Procedimento repetido até que satisfizesse a expressão 4.1	33
4.3	Modelo classificatório atuando no cenário “Levar compras” do FIM. a) Fluxo da aplicação sem efeito colateral indesejável. b) Fluxo da aplicação com efeito colateral indesejável.	34

LISTA DE TABELAS

2.1	Conjunto de objetos: seus atributos e classe pertencente. Adaptado de (Kotsiantis, 2007)	13
2.2	Medidas <i>recall</i> mensuradas a partir da aplicação da técnica <i>stratified-10-fold-cross-validation</i> com o classificador MLP sobre um dos <i>datasets</i> gerados nos testes iniciais deste trabalho. Tabela baseada em (Kerr et al., 2002).	17
4.1	Valores das curvas “ma x sl” da Figura 4.1 a partir do <i>dataset</i> com 40 instâncias. A (<i>Accuracy</i>), P (<i>Precision</i>), R (<i>Recall</i>), DP (Desvio Padrão), i (instâncias). Os valores A, P, R e DV estão em %.	33
4.2	Valores da validação do modelo final. DP (Desvio Padrão).	33

Introdução sobre do que se trata esta monografia e a maneira como o texto está organizado.

INTRODUÇÃO

A Internet nesta última década tem contribuído de forma significativa na economia e sociedade, deixando como legado uma sólida infraestrutura de rede de comunicação. O seu maior disseminador nesse período vem sendo *World Wide Web* (WWW), o qual permite o compartilhamento de informação e mídia de forma global (Chandrakanth et al., 2014).

A Internet está se tornando cada vez mais presente no cotidiano, devido, por exemplo, ao crescente número de usuários de dispositivos móveis, os quais possuem tecnologias de conexão com a Internet, as quais cada dia tornam-se mais acessíveis (Chandrakanth et al., 2014).

Em 2010 havia aproximadamente 1,5 bilhão de PCs conectados à Internet e mais que 1 bilhão de telefones móveis (Sundmaeker et al., 2010). Segundo Gartner¹, 6,4 bilhões de coisas estarão conectadas até o final de 2016 e, de acordo com estimativas da CISCO², em 2020 esse número atingirá cerca de 50 bilhões. A previsão de (Sundmaeker et al., 2010), a qual dizia que a denominada Internet dos PCs seria movida para o que se chama de Internet das Coisas fica então mais evidente neste atual cenário.

Internet of Things (IoT), traduzido para o português como Internet das Coisas, pode ser definida como uma rede de coisas que têm a capacidade de sentir, identificar e compreender o ambiente em que estão imersas (Ashton, 2009). “Coisas ou objetos”, tais como, RFID tags, sensores, telefones móveis, dentre outros, através de esquemas de endereçamento único são capazes de interagir com os outros e cooperar com seus vizinhos para alcançar um objetivo em comum (Atzori et al., 2010).

Uma definição tecnológica para IoT pode ser como em (Sundmaeker et al., 2010), o qual diz que IoT é parte integrante da futura Internet e pode ser definida como uma infraestrutura de rede global dinâmica com capacidades de autoconfiguração baseada nos padrões e interoperabilidade dos protocolos de comunicação onde coisas físicas e virtuais

¹<http://www.gartner.com/newsroom/id/3165317>

²http://www.cisco.com/c/en/us/solutions/collateral/enterprise/cisco-on-cisco/Cisco_IT_Trends_IoE_Is_the_New_Economy.html

têm identidade, atributos físicos, personalidade virtual, usam interfaces inteligentes e, são integradas dentro da rede de informações.

Apesar da grande potencialidade da Internet das Coisas, a qual gerou em 2015 um faturamento em torno de 130,33 bilhões de dólares americanos e tem prospecção de chegar até 883.55 bilhões em 2020³, ainda existem muitos desafios a serem vencidos, tais como segurança da informação (Roman et al., 2013), armazenamento e processamento de grande quantidade de dados (Zaslavsky et al., 2013), dentre outros. Adentrando em um dos desafios da IoT, o da disponibilidade de uma interface de comunicação e programação comum aos objetos, pode-se disponibilizar os dispositivos como serviços Web RESTFull, desta forma pode-se utilizar os protocolos Web como linguagem comum de integração dos dispositivos à Internet (Franca et al., 2011; Want et al., 2015; Mineraud et al., 2016).

Segundo (Heffelfinger, 2014), serviços Web RESTFull são serviços que seguem os princípios arquiteturais REST (Representational State Transfer), onde os serviços Web são vistos como recursos e podem ser identificados por Uniform Resources Identifiers (URIs).

Os principais princípios do estilo arquitetural REST são: endereçamento global através de identificação dos recursos, interface uniforme compartilhada por todos os recursos, interações *stateless* entre os serviços e, mensagens auto-descritivas e *hypermedia* como um mecanismo para descoberta descentralizada de recursos por referência (Pautasso, 2014).

Diante deste cenário onde diversos dispositivos são combinados para alcançar um objetivo em comum (Kranz et al., 2010; Atzori et al., 2010; Whitmore et al., 2015) e pelo fato destes comumente serem disponibilizados como serviços Web, e da crescente adição destes à IoT, surge a necessidade de tratar um outro problema ainda não citado, que ocorre quando a composição de dispositivos (os quais oferecem alguma funcionalidade) leva a um comportamento não esperado (interação de características) nesta composição, o qual pode resultar em um estado inconsistente, instabilidade ou dados imprecisos - efeito colateral indesejável (NHLABATSI et al., 2008).

Dentre as possíveis causas (Weiss et al., 2007) de efeitos colaterais indesejáveis pode-se citar:

- *Goal Conflict* (Conflito de interesses): Cada característica ou unidade de adicional funcionalidade ao software (funcionalidade) tem seus próprios objetivos (designadas para fazer algum tipo de processamento, coletar algum dado, dar uma saída específica, dentre outros) a serem alcançados. Entretanto, quando as características são combinadas, os objetivos das duas características podem entrar em conflito e não se pode garantir que todos sejam alcançados.
- *Violation of Assumptions* (Violação de suposições): Desenvolvedores das funcionalidades de software precisam fazer algumas suposições de como outras características trabalham. Estes podem fazer suposições incorretas, por exemplo, devido a semântica ambígua (tal como uso do mesmo conceito de diferentes formas), ou devido a presença de diferentes versões da mesma funcionalidade de software. De

³<http://www.marketsandmarkets.com/Market-Reports/iot-application-technology-market-258239167.html>

forma similar, implementação de características podem ser baseadas em suposições incorretas sobre o contexto de uso. Uma das características de uma *Violation of Assumptions* é quando a mudança em uma *feature* quebra a suposição correta que a outra característica tinha sobre esta.

- *Policy Conflict* (Conflito de políticas): Políticas proveem os meios para especificação e modularização do comportamento de uma característica, na medida em que alinha suas capacidades e restrições com os requisitos de seus usuários. A *Policy Conflict* ocorre caso exista políticas (autenticação, ou de privacidade) especificadas em duas *features* que se referem as suas correspondentes operações, e que as políticas não são compatíveis.

O presente trabalho propõe:

- Criar um cenário (“Levar compras”) que simule efeitos colaterais indesejáveis diante da visão da IoT, os quais são causados por violação de hipótese. Para isto, é realizado um estudo de caso em um *Home Network System* (HNS). HNS é uma rede doméstica de coisas (aparelhos domésticos e sensores) com capacidade de conectividade de rede e, interface de controle e monitoramento. Nesse tipo de sistema os aparelhos e sensores podem juntos prover novas funcionalidades, as quais atendam as expectativas do usuário da casa. O HNS tem um componente denominado de *Home Server* (HS), este acessa os dispositivos através de APIs e disponibiliza as funcionalidades ao usuário final (Nakamura et al., 2009; Nakamura et al., 2013). Os dispositivos no cenário “Levar compras” são serviços independentes disponibilizados como serviços Web RESTFul.
- Detectar efeitos colaterais indesejáveis de forma inteligente no cenário “Levar compras” utilizando o método *ensemble* DECORATE. Este tipo de método utiliza um conjunto de classificadores para construir diversas hipóteses (modelos). Então é utilizado o conhecimento de cada modelo afim de se chegar em uma decisão final mais confiável (Melville e Mooney, 2004). Diversas propostas de detecção de efeitos colaterais indesejáveis têm sido abordadas em HNS (Wilson et al., 2008; Nakamura et al., 2009; Nakamura et al., 2013; Maternaghan e Turner, 2013; Alfakeeh e Al-Bayatti, 2016), nenhuma destas propostas citadas utiliza alguma metodologia de detecção que utilize classificadores para detecção de efeitos colaterais indesejáveis. Na arquitetura de HNS utilizada nestes trabalhos, os dispositivos são utilizados para prover serviços ao usuário da casa, e estes são acessados somente pelo HS através das APIs dos dispositivos. Os dispositivos nesta arquitetura não têm capacidade de se comunicar e oferecer serviços entre eles de forma independente. No trabalho desenvolvido neste TCC, os dispositivos (elevador, garra e veículo terrestre) são disponibilizados como serviços Web RESTFul e estes têm capacidade de se comunicar e oferecer serviços de forma independente, representando desta forma um cenário da IoT.

O cenário “Levar compras” é a composição dos três dispositivos independentes citados (elevador, garra e veículo terrestre) os quais juntos compõem a funcionalidade “Levar

compras”, que é disponibilizada ao usuário humano do HNS através do HS (que também é um serviço Web RESTFul).

Através do *ensemble* DECORATE foi possível detectar os efeitos colaterais indesejáveis com mais de 90% de precisão.

O restante deste trabalho está organizado da seguinte maneira:

- Capítulo 2: Apresenta todo embasamento teórico necessário para compreender o estudo realizado neste trabalho.
- Capítulo 3: Esse capítulo descreve a proposta para detecção de efeitos colaterais indesejáveis no cenário “Levar compras”.
- Capítulo 4: Trata dos procedimentos realizados para validar a proposta de pesquisa deste TCC.
- Capítulo 5: Descreve de forma resumida os trabalhos relacionados ao desta monografia, assim como em que difere destes.
- Capítulo 6: Este capítulo é uma síntese da investigação e dos experimentos realizados neste TCC.

Este capítulo tem como objetivo fundamentar as bases necessárias dos campos de estudos utilizados nesta monografia.

FUNDAMENTAÇÃO TEÓRICA

2.1 INTERNET DAS COISAS

Há anos atrás, por volta de 2009, a informação gerada na Internet era quase toda dependente da ação humana. Os aproximadamente 50 PB *petabytes* gerados na Internet até este ano de 2009 foram gerados por humanos. Como, por exemplo, através de um clique numa câmera digital, digitalização de um código de barras, digitando texto, pressionando um botão para gravar. O problema é que o ser humano tem capacidade finita de atenção e acurácia, o que não é bom para capturar dados das coisas do mundo real. Sendo assim, há a necessidade de prover aos computadores seus próprios meios de coletar informação para que estes possam sentir o ambiente em que se encontram. RFID e sensores permitem que estes observem, identifiquem e compreendam o ambiente em que estão imersos sem a necessidade de dados que são introduzidos por humanos (Ashton, 2009). Assim sendo, *Internet of Things* (IoT), traduzido para o português como Internet das Coisas, pode ser definida como uma rede de coisas que têm a capacidade de sentir, identificar e compreender o ambiente em que estão imersas.

“Coisas ou objetos”, tais como, RFID tags, sensores, telefones móveis, dentre outros, através de esquemas de endereçamento único são capazes de interagir com os outros e cooperar com seus vizinhos para alcançar um objetivo em comum (Atzori et al., 2010). Outros exemplos de “coisas ou objetos” podem ser pessoas, geladeiras, televisores, veículos, roupas, medicações, livros, passaportes, contanto que possam ser identificadas unicamente e possam se comunicar com as outras coisas e/ou possam ser acessados remotamente por humanos ou por máquinas.

Uma definição tecnológica para IoT pode ser como em (Sundmaeker et al., 2010), o qual diz que IoT é parte integrante da futura Internet e pode ser definida como uma infraestrutura de rede global dinâmica com capacidades de autoconfiguração baseada nos padrões e interoperabilidade dos protocolos de comunicação onde coisas físicas e virtuais têm identidade, atributos físicos, personalidade virtual, usam interfaces inteligentes e, são integradas dentro da rede de informações.

Diante deste cenário pode-se citar exemplos de aplicações na IoT para diversos domínios, tais como, logística e transporte; cuidados com a saúde; ambiente inteligente (casa (Seção 2.3), escritório); (Atzori et al., 2010) verificação de procedência alimentícia; dentre outros. Seguem dois exemplos de aplicações, uma na área de cuidados com a saúde e outra na verificação de procedência de alimentos, respectivamente.

- Dispositivos implantáveis com capacidade de comunicação sem fio podem ser utilizados para armazenar registros sobre a saúde de um paciente em situações de risco e podem ser decisivos para salvar a vida do paciente. A capacidade de ter acesso a essas informações nessas circunstâncias fazem com que hospitais possam saber de imediato como tratar um paciente que esta a caminho. Esta possibilidade é especialmente útil para pessoas com diabetes, câncer, problemas de coração na artéria coronária, doenças do pulmão, assim como as pessoas com implantes médicos complexos, tais como, marcapassos, tubos, transplantes de órgãos e aqueles que podem ficar inconscientes e incapazes de comunicar-se durante uma operação (Weber e Weber, 2010).
- Rastreabilidade de produtos alimentícios ajudam os usuários a verificar a origem de um produto, assim como informações de composição química, dentre outros. Mas também previne de doenças não desejadas. Por exemplo, avisos atuais sobre a mercadoria em questão podem ser disponibilizados à medida que o produto sai da origem e vai passando para os outros níveis de consumo, desta forma os consumidores podem evitar o contágio de doenças como gripe aviária e, encefalopatia espongiforme bovina (EEB), mais conhecida como doença da vaca louca (Weber e Weber, 2010).

2.1.1 Dispositivos como serviços Web RESTful

Serviços Web surgiram tendo como principal foco o reúso de aplicações existentes (dentre as quais incluíam código fonte legado) para que pudessem se integrar de forma leve com outras aplicações. Geralmente essa integração tinha como referência o desejo de novas formas de compartilhamento dos serviços ao longo das diversas linhas do negócio ou entre parceiros (Papazoglou, 2008).

Um dos desafios referentes a visão da IoT é sua interoperabilidade e, como possível solução se pode disponibilizar os dispositivos como serviços Web, por exemplo, seguindo os princípios REST. Desta maneira os dispositivos podem interagir entre si e com outros sistemas na Web.

Segundo (Heffelfinger, 2014), Representational State Transfer (REST) é um estilo de arquitetura no qual serviços Web são vistos como recursos e podem ser identificados por Uniform Resources Identifiers (URIs). Serviços Web que são desenvolvidos usando REST são conhecidos como RESTful web services.

Os principais princípios do estilo arquitetural REST são: endereçamento global através de identificação dos recursos, interface uniforme compartilhada por todos os recursos, interações *stateless* entre os serviços e, mensagens auto-descritivas e *hypermedia* como um

mecanismo para descoberta descentralizada de recursos por referência (Pautasso, 2014), conforme abaixo.

1. Identificação dos recursos - todos os recursos que são publicados por um serviço Web deveriam ser disponibilizados por um único e estável identificador global (Pautasso, 2014). A exemplo das URIs (Franca et al., 2011).
2. Interface uniforme - todos os recursos interagem através de uma interface uniforme, a qual prover um conjunto de métodos pequeno, genérico e funcionalmente suficiente para suportar todas as possíveis interações entre os serviços. Cada método tem uma semântica bem definida em relação aos efeitos que causará no estado do recurso. No contexto da Web e do protocolo HTTP que é utilizado, “Interface uniforme” pode ser alcançado com os seus métodos (GET, PUT, DELETE, POST, HEAD, OPTIONS, dentre outros) os quais podem ser aplicados para todos os identificadores dos recursos Web (URIs) (Pautasso, 2014).
3. Interações *stateless* - os serviços não podem estabelecer sessões permanentes entre eles. Isto assegura que as requisições para um recurso sejam independentes entre si. No final de cada interação, não há estados compartilhados entre clientes e servidores. Requisições podem resultar em uma mudança de estado do recurso, mas o novo estado é imediatamente visível para todos clientes (Pautasso, 2014).
4. Mensagens auto-descritivas - Serviços interagem através de requisição e mensagem de resposta, que contem tanto os dados (representações dos recursos) e correspondente *meta-data*. As representações podem variar de acordo com o contexto do cliente, interesses e habilidades. Por exemplo, um cliente *mobile* pode obter uma representação do recurso que exige pouco consumo de banda de dados. Da mesma forma, um navegador pode requisitar a representação de uma página Web em uma linguagem específica, de acordo com suas preferências. Esta característica aumenta de maneira significativa o grau de interoperabilidade, pois um cliente pode dinamicamente negociar o mais apropriado formato de representação com o recurso (também conhecido como *media type*), em vez de ser forçado a sempre trabalhar com uma determinada representação do recurso. As requisições e mensagens de respostas também devem conter explicitamente *meta-data* sobre sua representação, desta maneira os serviços não precisam assumir algum tipo de acordo de sobrecarga no sentido de como tal representação seria analisada, processada e entendida (Pautasso, 2014).
5. *Hypermedia* - Recursos podem ser relacionados entre si. *Hypermedia* embute referências a outros recursos relacionados ou dentro de suas representações ou em sua correspondente *metadata*. Clientes então podem descobrir os *hyper-links* dos recursos relacionados ao processar suas representações e escolher seguir o link. Como exemplificado em (Franca et al., 2011), um sistema de uma instituição que possui um recurso *lista_cursos* (lista todos os cursos da instituição), este pode oferecer os *hyper-links* que representam cada recurso que representa um determinado curso.

Hypermedia auxilia na descoberta descentralizada de recursos e também pode ser utilizada para descoberta e descrição de protocolos de interação entre os serviços. Pelo fato deste ser o princípio menos utilizado pelas APIs que se dizem ser REST-Ful, algumas vezes as APIs disponibilizadas por serviços Web que além das outras restrições contemplam esta, são também chamadas de *Hypermedia APIs* (Pautasso, 2014).

A disponibilização dos dispositivos como serviços Web RESTful vêm sendo empregada através de duas abordagens. Na primeira, quando os dispositivos têm recursos suficientes (memória, processamento e largura de banda de rede) servidores Web são embarcados nos próprios dispositivos e estes são disponibilizados como recursos REST. Enquanto que na segunda, quando uma coisa não possui recursos suficientes para tal propósito, utiliza-se de outro dispositivo, por exemplo, um ¹Raspberry Pi ou qualquer outro controlador com recursos suficientes para instalação e execução de um servidor Web. Este então atua como um intermediador entre o objeto e a Internet (Franca et al., 2011).

Adotando esse padrão os dispositivos podem ter suas propriedades disponíveis através de qualquer navegador, sem a necessidade de instalação de nenhum programa ou driver adicional (Want et al., 2015), como pode ser exemplificado na Figura 2.1. Nesta, o dispositivo tem suas informações acessadas através de uma requisição GET, que é realizada passando o URI (endereço que identifica a funcionalidade do serviço de forma única) do dispositivo ao navegador, e obtêm-se um retorno no formato JSON², o qual tem as informações do dispositivo: ligado (informa se o dispositivo está em operação); obstáculo (informa se existe algum obstáculo na linha do sensor); anguloDetecção (informa o ângulo de detecção); faixa de detecção (informa qual a faixa de detecção ajustada no momento); status (4 - informa que a requisição ao dispositivo foi processada sem problemas).



Figura 2.1: Dispositivo sensor de obstáculo do elevador disponibilizado como serviço. Faixa de detecção compatível com a largura do elevador utilizado na maquete do cenário (Figura 3.3)

¹<https://www.raspberrypi.org/products/>

²Formato de dados leve para troca de mensagens. (<http://www.json.org/>)

Além disso, mashups físicos³ podem ser construídos com muito menos esforço do que as existentes abordagens, minimizando a barreira para o desenvolvimento de aplicações com dispositivos (Guinard e Trifa, 2009). Pelo fato das coisas serem disponibilizadas como serviços Web, ainda pode-se utilizar-se dos outros recursos disponíveis na Web, a exemplo de sistemas de cache, balanceamento de carga, indexação e pesquisa (Franca et al., 2011). Assim então promovendo a visão da Internet das Coisas.

Diversas comparações (Mulligan e Gracanin, 2009; Castillo et al., 2011; Belqasmi et al., 2012; Wagh e Thool, 2012) entre serviços Web baseados em SOAP e REST foram realizadas e, devido a natureza de muitos dispositivos da IoT, os quais possuem poucos recursos de hardware, os princípios REST têm sido amplamente aplicados na integração dos dispositivos inteligentes a Web (Franca et al., 2011).

Diante do cenário da IoT descrito até agora faz-se observar que muitos objetos quando isolados, não tem como prover funcionalidade(s) que atendam ao(s) requisito(s) de um usuário final (outra coisa ou um humano). Então, para atender a tais requisitos, há necessidade de compor estes objetos com outros para oferecer nova(s) funcionalidade(s) que atendam aos usuários finais. Entretanto, a composição das coisas pode levar a uma interação de características (assunto abordado na próxima Seção 2.2) a qual pode provocar efeitos colaterais indesejáveis.

2.2 INTERAÇÃO DE CARACTERÍSTICAS

Em desenvolvimento de *software*, uma *feature* (característica) é um componente de adicional funcionalidade ao *software* (Calder et al., 2003), consistindo de um conjunto de requisitos logicamente relacionados e suas especificações, o qual se destina a fornecer um determinado efeito comportamental (NHLABATSI et al., 2008).

Poucas *features* atendem aos requisitos do usuário final quando estão isoladas. Para resolver este problema, características são combinadas para fornecer um novo componente de funcionalidade ao software. Entretanto, quando a composição de *features* leva a algum comportamento não esperado - interação de características, esta pode resultar em efeitos colaterais indesejáveis (NHLABATSI et al., 2008) ou em efeitos colaterais desejáveis (Weiss et al., 2005; Wilson et al., 2005).

Efeito colateral indesejável é quando há interação de características e esta resulta em um estado inconsistente do sistema, um sistema instável ou dados imprecisos (NHLABATSI et al., 2008). Já efeito colateral desejável é quando a interação de características resulta num comportamento que ajudará de algum modo a funcionalidade em questão (Weiss et al., 2005; Wilson et al., 2005).

Existem diversas causas para interação de características, algumas delas são listadas a seguir de forma mais genérica possível (Weiss et al., 2007).

- *Goal Conflict* (Conflito de interesses): Cada característica tem seus próprios objetivos (designadas para fazer algum tipo de processamento, coletar algum dado, dar uma saída específica, dentre outros) a serem alcançados. Entretanto, quando as

³Aplicativos criados a partir da composição de dados e serviços de dispositivos físicos com outros recursos Web.

características são combinadas, os objetivos das duas características podem entrar em conflito e não se pode garantir que todos sejam alcançados.

- *Resource Contention* (Contenção de recurso): *Features* podem estar competindo umas com as outras devido ao acesso a recursos de capacidade limitada, tais como, memória, CPU, largura de banda, acesso a banco de dados, dentre outros. Desta forma, a correta execução de uma característica pode ser comprometida por outra característica que esteja utilizando além da sua cota de um recurso compartilhado.
- *Violation of Assumptions* (Violação de suposições): Desenvolvedores das funcionalidades de software precisam fazer algumas suposições de como outras características trabalham. Estes podem fazer suposições incorretas, por exemplo, devido a semântica ambígua (tal como uso do mesmo conceito de diferentes formas), ou devido a presença de diferentes versões da mesma funcionalidade de software. De forma similar, implementação de características podem ser baseadas em suposições incorretas sobre o contexto de uso. Uma das características de uma *Violation of Assumptions* é quando a mudança em uma *feature* quebra a suposição correta que a outra característica tinha sobre esta.
- *Policy Conflict* (Conflito de políticas): Políticas proveem os meios para especificação e modularização do comportamento de uma característica, na medida em que alinha suas capacidades e restrições com os requisitos de seus usuários. A *Policy Conflict* ocorre caso exista políticas (autenticação, ou de privacidade) especificadas em duas *features* que se referem as suas correspondentes operações, e que as políticas não são compatíveis.

Diversas taxonomias foram propostas para descrever os tipos de interação de características. Algumas destas propostas foram produzidas por Cameron e Velthuijsen (domínio das telecomunicações), Kolberg (domínio das *smarthomes*), e Shehata. Este último fez uma síntese das duas taxonomias citadas para uma taxonomia genérica a qual consiste de nove tipos de interação de características. Esta taxonomia tem a pretensão de ser aplicável na maioria dos domínios (NHLABATSI et al., 2008).

A seguir é apresentada e explicada os tipos de interação de características propostos por Shehata em uma versão reduzida da sua taxonomia (NHLABATSI et al., 2008). Para as explicações considere pré-condições como sendo as condições necessárias para que uma característica inicie sua execução e, como pós-condições as ações ou saídas esperadas após o término da execução da característica.

- *Dependence*: A execução de uma característica depende da execução correta da outra.
- *Non-determinism*: Ocorre quando características têm as mesmas pré-condições, mas pós-condições diferentes, ou seja, duas ou mais especificações de requisitos requerem que um domínio compartilhado tenha comportamentos distintos simultaneamente, quando o domínio pode somente ter um comportamento por vez. Domínio aqui significa uma propriedade do ambiente a qual uma especificação de uma característica utiliza para satisfazer seus requerimentos.

- *Negative impact*: Similar a *Non-determinism*, no sentido que as características sobreponham pré-condições. Mas diferente no sentido que ambas as *features* são executadas e o resultado das suas pós-condições são inconsistentes. As pós-condições de uma característica diminui o efeito das pós-condições da(s) outra(s).
- *Invocation Order*: O comportamento da execução das características em uma determinada sequência é diferente do comportamento destas quando há mudança na sequência.
- *Bypass*: Uma característica F1 *bypasses* a execução de outra característica F2 se F1 muda o estado do ambiente compartilhado de tal forma que o novo estado não atende as pré-condições da outra característica F2.
- *Infinite Looping*: Ocorre quando duas *features* são reciprocamente ligadas em suas pós-condições e disparos de eventos. As características F1 e F2 são reciprocamente ligadas se as pós-condições de F1 cria um disparo de evento para F2 e vice e versa. Suponha que F1 é disparada e inicia sua execução criando um disparo de evento para F2. F2 inicia sua execução e cria um disparo para F1. Este processo então se repete indefinidamente, criando um loop infinito.

Autores em (NHLABATSI et al., 2008) não consideram *dependence* como uma causa de interação de características, pois este diz que *dependence* passa uma fraca noção de interação de características, já que esta implica que uma característica não funciona corretamente isoladamente e, segundo os autores, uma importante propriedade é que a *feature* deva satisfazer seus requisitos quando isolada, mas quando em composição pode acontecer a quebra desses requisitos.

Dentro deste cenário de interação de características, diversos métodos vêm sendo propostos para detecção e/ou resolução dessas interações. Duas abordagens são adotadas para este fim: *Online techniques* (técnicas online) e *Offline techniques* (técnicas offline). Técnicas online são métodos aplicados enquanto o sistema está sendo executado dentro de uma rede, ou seja, as *features* já foram desenvolvidas e estão em fase de produção sendo executadas. São uma combinação de mecanismos de detecção e resolução de interação de características. Já as técnicas offline (abordagens da engenharia de software e métodos formais) são aquelas aplicadas durante a fase desenvolvimento das características (Calder et al., 2003).

Dentre as técnicas offline, a abordagem da engenharia de software vêm propondo diversos caminhos (Thum et al., 2014; Siegmund et al., 2012a; Siegmund et al., 2012b) para auxiliar na remoção de interação de características antes de sua implantação (Calder et al., 2003). Na abordagem dos métodos formais (Almeida et al., 2011) uma variedade de técnicas também têm sido adotadas, tais como, lógicas clássicas e construtivas, autômatos de estado finito e infinito, autômatos estendidos, *petri-nets*, sistemas de transição e, linguagens como SDL, Promela, Z e LOTOS (Calder et al., 2003).

Dentre as técnicas online, *Feature Interaction Manager - FIM* é uma das técnicas que vem sendo adotada em *Home Network Systems* ou *Smart homes* (Wilson et al., 2005; Wilson et al., 2008; Nakamura et al., 2009). FIM (Calder et al., 2003) é um componente

do sistema introduzido dentro de uma rede com a capacidade de observar e controlar as chamadas a qualquer *feature*. Este trabalho concentra-se nesta técnica para prover detecção de efeitos colaterais indesejáveis.

2.3 HOME NETWORK SYSTEM

Diferentes sensores e aparelhos domésticos a exemplo de lâmpadas, ar-condicionadores, sistema de segurança e entretenimento, além de interagirem dentro do ambiente doméstico, podem ser conectados a Internet e controlados por um *smartphone* ou qualquer outro dispositivo da IoT. Desta forma pode-se não só controlar os dispositivos como também monitorar o ambiente doméstico (temperatura, consumo de energia, dentre outros). Uma casa com essas características é comumente chamada de *Smart Home* e é normalmente implementada como um *Home Network System* (Piyare, 2013).

Um *Home Network System* (HNS) é uma rede doméstica de coisas (aparelhos domésticos e sensores) com capacidade de conectividade de rede e, interface de controle e monitoramento. Nesse tipo de sistema os aparelhos e sensores podem juntos prover novas funcionalidades, as quais atendam as expectativas do usuário, como por exemplo, “Levar compras” (Seção 3.2). Os aparelhos e sensores dessa rede podem ser disponibilizados como serviços Web RESTful. O HNS tem um componente denominado de *Home Server* (HS), este acessa os dispositivos através de APIs e disponibiliza as funcionalidades ao usuário final. O HS também pode atuar de outra forma, detectando e resolvendo efeitos colaterais indesejáveis, neste caso o HS pode ser chamado de *Feature Interaction Manager* - FIM (Seção 2.2). O HS ainda pode servir de mediador para redes externas, ou seja, este pode disponibilizar cada serviço de forma única na Internet, para isto, basta ter um endereço IP público estático e oferecer seus serviços através de uma API, por exemplo, uma API que segue os princípios REST. Desta forma, cada dispositivo do HNS pode ser identificado unicamente na chamada Internet das Coisas (Nakamura et al., 2009), (Nakamura et al., 2013). Este trabalho replica um cenário de HNS, conforme descrito na Seção 3.2.

Para que o FIM (Seção 2.2) possa prover a detecção de efeitos colaterais indesejáveis é proposto (Capítulo 3) utilizar um procedimento classificatório a fim de se obter um modelo final satisfatório, tema abordado na Seção 2.4 seguinte.

2.4 CLASSIFICAÇÃO - APRENDIZADO SUPERVISIONADO

Em seu sentido mais geral, o termo classificação poderia cobrir qualquer contexto no qual alguma decisão ou predição é realizada baseada nas informações presentes. Classificação então pode ser caracterizada como um método formal que realiza julgamentos repetidos a cada nova situação. No caso deste trabalho, o termo classificação será restringido a situação de se construir um procedimento classificatório com base em um conjunto de dados no qual as classes são conhecidas, comumente chamado de aprendizado supervisionado (Michie et al., 1994). O objetivo final de se construir esse procedimento classificatório é que este possa generalizar para conjuntos de dados que não participaram da sua construção, ou seja, o classificador tem que ser capaz de classificar corretamente

novos dados (Kotsiantis, 2007). O conjunto de dados aqui citado é composto de objetos (instâncias), sendo cada instância representada por um conjunto de atributos e, cada uma é pertencente a uma classe específica (ver Tabela 2.1).

Objeto	atributo 1	atributo 2	..	atributo n	classe
1	10	3		4	A
2	11	11		9	A
3	14	26		6	B
..				..	

Tabela 2.1: Conjunto de objetos: seus atributos e classe pertencente. Adaptado de (Kotsiantis, 2007)

Muitos problemas nas áreas da ciência, indústria, medicina, dentre outros, podem ser tratados como problemas de classificação. A exemplo de diagnósticos médicos (classificar tumores como benigno ou maligno), classificação de transações de cartão de crédito como legítima ou fraudulenta, controle de qualidade, reconhecimento de fala (Zhang, 2000).

De modo geral, a metodologia adotada para construção de um classificador é como segue e pode ser visualizado na Figura 2.2.

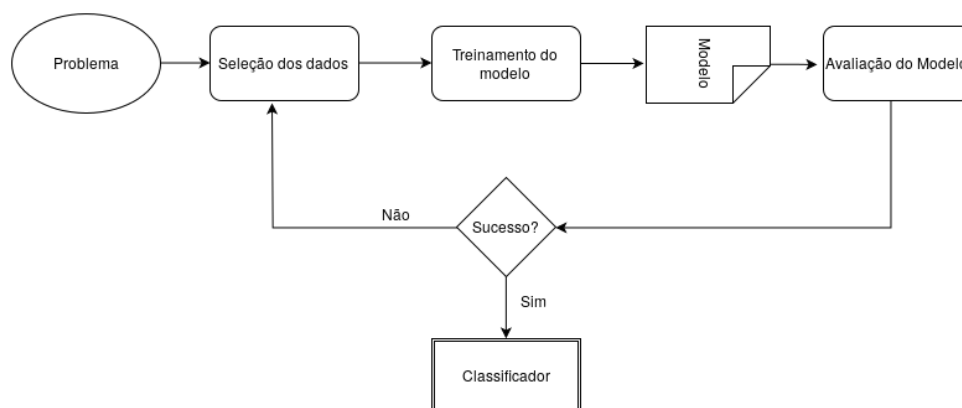


Figura 2.2: Processo de construção de um classificador em aprendizado supervisionado. Baseado em (Kotsiantis, 2007) e proft⁴.

1. Problema: primeiramente se deve saber qual problema quer resolver.
2. Seleção dos dados: coleta do conjunto de dados (*dataset*). Se existir uma pessoa com alto conhecimento sobre os atributos, este pode sugerir quais atributos são mais significativos para o problema. Caso contrário o mais simples método é mensurar todos os atributos na esperança que as características corretas possam ser isoladas (Kotsiantis, 2007).

⁴<http://en.proft.me/2015/12/24/types-machine-learning-algorithms>

3. Treinamento do modelo: treina-se o dataset da etapa anterior sobre um algoritmo de classificação (árvores de decisão, *Fisher's linear discriminants*, redes neurais - *Multi Layer Perceptron (MLP)* (Michie et al., 1994)). Então obtêm-se um modelo, que é o algoritmo de classificação treinado com o dataset.
4. Avaliação do Modelo: é baseada em métodos avaliativos (ver Seção 2.4.2), os quais têm o propósito de verificar a generalização do modelo, ou seja, verificar o quão bom é o modelo do classificador quando submetido a instâncias desconhecidas (que não fazem parte do modelo). Uma das medidas utilizadas por tais métodos é a taxa de acurácia (percentagem das predições feitas corretamente dividida pelo número total de predições realizadas) (Kotsiantis, 2007).
5. Classificador: se o modelo foi validado na etapa anterior, este generaliza o suficiente para o problema proposto, então este modelo pode ser utilizado como o classificador final (pode ser implantado em um sistema), caso contrário reinicia-se o processo a partir da “seleção de dados” reavaliando as decisões de cada etapa.

Dentre os algoritmos citados na etapa de treinamento da Figura 2.2, as redes neurais são modelos que atendem bem tanto problemas de classificação que não são linearmente separáveis (problemas que se adequam mais a realidade), quanto problemas que são linearmente separáveis, entretanto, utilizar-se das redes neurais para problemas linearmente separáveis é totalmente desaconselhável, já que pode ser muito mais custoso computacionalmente. (Zhang, 2000) (Elizondo, 2006) A Figura 2.3 ilustra problema linearmente separável e problema não separável linearmente. Apesar de alguns problemas serem não linearmente separáveis, estes podem ser separados de forma linear sem perdas substanciais (Figura 3.9).

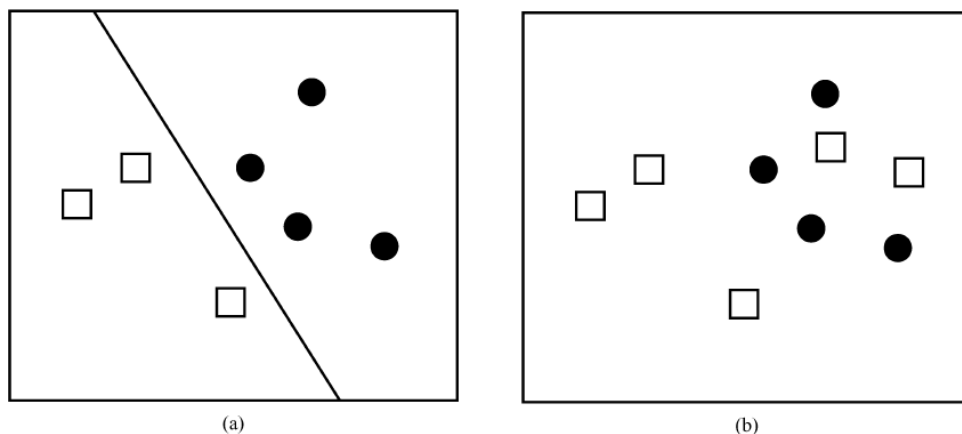


Figura 2.3: (a) Problema linearmente separável. (b) Problema não separável linearmente. (Elizondo, 2006)

Algumas comparações (Firdausi et al., 2010; Arora e Suman, 2012; Karthikeyan e Thangaraju, 2013) vêm sendo realizadas entre J48 (árvore de decisão, implementação WEKA (Hall et al., 2009) do algoritmo C4.5 (Quinlan, 1993)) e MLP (rede neural). Nessas, J48 obteve resultados melhores ou próximos para problemas de classificação binária

(entre duas classes) e mostrou-se ser muito menos custoso computacionalmente. Sendo assim, o uso do algoritmo J48 parece ser uma boa escolha e, este será utilizado como base do *DECORATE*, ver Seção 2.4.1, o qual será o método de classificação utilizado neste trabalho.

2.4.1 DECORATE - um método ensemble

Um método *ensemble* combina as decisões de diversos classificadores para chegar numa decisão final. A diversidade das hipóteses dos classificadores do *ensemble* é tida como um importante fator para obter uma boa generalização. O DECORATE (*Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples*) constrói diversas hipóteses utilizando datasets de treino adicionais gerados artificialmente. Resultados experimentais demonstraram que o DECORATE tendo como algoritmo base o J48 conseguiu obter taxas de acurácia satisfatórias, mais altas que outros métodos *ensemble* e, melhores do que se o J48 fosse utilizado de forma isolada. O DECORATE também se saiu muito bem em *datasets* pequenos, nos quais obteve um bom grau de generalização (Melville e Mooney, 2003; Melville e Mooney, 2004).

No DECORATE, o *ensemble* é gerado de forma iterativa, ou seja, a cada iteração é gerado um classificador e este é adicionado ao *ensemble*. Na primeira iteração o *ensemble* contém o classificador treinado no *dataset* original. A partir da segunda iteração os classificadores são gerados com o *dataset* original mais dados artificiais. O número de exemplos artificiais a serem gerados é especificado como um fator em cima do tamanho do *dataset* de treino. Caso a adição do novo classificador ao *ensemble* aumente sua taxa de erro, este é ignorado, senão é adicionado ao *ensemble*. Este procedimento é repetido até que se alcance um número de *committee* (membro do comitê julgador ou classificador) desejado ou exceda o limite de iterações (Melville e Mooney, 2003).

Para classificar um exemplo não rotulado, x , utiliza-se a seguinte fórmula 2.1, onde C_i é um classificador pertencente ao *ensemble* C^* , $|C^*|$ representa a quantidade de classificadores no *ensemble* e $P_{C_i,y}(x)$ é a probabilidade de x pertencer a classe y de acordo com o classificador C_i .

$$P_y(x) = \frac{\sum_{C_i \in C^*} P_{C_i,y}(x)}{|C^*|} \quad (2.1)$$

2.4.2 Métodos avaliativos

Métodos avaliativos têm o propósito de verificar a generalização do modelo, ou seja, verificar quão bom é o modelo do classificador quando submetido a instâncias desconhecidas (que não fazem parte do modelo). Para isto, utiliza-se de medidas estatísticas para auxiliar na tarefa (Witten e Frank, 2005). Como por exemplo, acurácia, *precision*, *recall*, dentre outras. As medidas estatísticas citadas são baseadas em quatro componentes: verdadeiro positivo (*True Positive* - *TP*), verdadeiro negativo (*True Negative* - *TN*), falso positivo (*False Positive* - *FP*) e falso negativo (*False Negative* - *FN*) (Davis e Goadrich, 2006).

- Verdadeiro positivo (TP): instância que pertence a classe positiva e que foi corretamente classificada como positiva.
- Verdadeiro negativo (TN): instância que pertence a classe negativa e que foi classificada corretamente como negativa.
- Falso positivo (FP): instância que pertence a classe negativa e que foi classificada incorretamente como positiva.
- Falso negativo (FN): instância que pertence a classe positiva e que foi classificada incorretamente como negativa.
- *Accuracy*: taxa correta de classificação em relação a todos os exemplos, a qual pode ser visualizada e obtida através da fórmula 2.2 (Metz, 1978).
- *Precision*: mede a taxa de exemplos classificados como positivos que realmente são positivos. Ou seja, dentre todos os exemplos classificados como positivos, quais realmente são positivos? Esta relação pode ser visualizada e obtida através da fórmula 2.3 (Davis e Goadrich, 2006).
- *Recall* ou *True Positive Rate* - *TPR*: mede a taxa de exemplos positivos que foram corretamente classificados. Ou seja, dentre os exemplos classificados como positivos (que realmente são positivos) e os classificados como negativos (mas que são positivos), qual a taxa de exemplos positivos classificados corretamente? Esta relação pode ser visualizada e obtida através da fórmula 2.4 (Davis e Goadrich, 2006).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.2)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.3)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.4)$$

Outra medida utilizada é o desvio padrão, utilizada para verificar a dispersão de um conjunto de dados medido, por exemplo, o desvio padrão de cada conjunto das medidas avaliativas citadas anteriormente (*Precision*, *Recall*, *Accuracy*) quando aplicado o *stratified-k-fold-cross-validation*. Um desvio padrão baixo indica que os elementos do conjunto têm valores próximos ao esperado, enquanto que valores altos indicam que os elementos estão dispersos dentro de uma faixa grande de valores (Kerr et al., 2002), (Bland e Altman, 1996).

O desvio padrão pode ser medido de acordo com a fórmula 2.5, que é a $\sqrt{\text{variância}}$. Onde x_i é um elemento do conjunto, \bar{x} é a média aritmética dos i elementos, e n é a quantidade de elementos do conjunto. Usualmente o desvio padrão é também calculado para somente um subconjunto das medidas, desta forma o valor produzido é uma estimativa do desvio padrão para a população dos dados em questão. Para este caso em específico,

a fórmula 2.5 sofre uma pequena alteração (fórmula 2.6, ao invés de n , tem-se $n-1$), a fim de corrigir os possíveis erros gerados por tal estimativa (Kerr et al., 2002).

$$\text{Desvio padrão } (S) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad (2.5)$$

$$\text{Desvio padrão } (S) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (2.6)$$

Uma exemplificação da mensuração do desvio padrão pode ser realizada com os dados da Tabela 2.2 e aplicação da fórmula 2.5. Os elementos x_i (*recall*) da população foram gerados utilizando *stratified-10-fold-cross-validation* em cima de um dos *datasets* gerados através de testes iniciais realizados neste trabalho com o classificador MLP. A soma dos $x_i - \bar{x}$, na segunda coluna da tabela, gerou um valor diferente de zero devido a aproximação realizada para duas casas decimais, o correto seria 0 (zero). A aplicação da fórmula neste cenário pode ser visualizada em 2.7.

$$\text{Desvio padrão } (S) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{4173.42}{10}} = 24.43\% \quad (2.7)$$

$x_i(\%)$	$x_i - \bar{x}(\%)$	$(x_i - \bar{x})^2(\%)^2$
100.00	19.17	367.49
66.67	14.16	200.50
100.00	19.17	367.49
100.00	19.17	367.49
75.00	5.83	33.99
100.00	19.17	367.49
66.67	14.16	200.50
50.00	30.83	950.49
50.00	30.83	950.49
100.00	19.17	367.49
$\sum = 808.34$	0.04	4173.42

Tabela 2.2: Medidas *recall* mensuradas a partir da aplicação da técnica *stratified-10-fold-cross-validation* com o classificador MLP sobre um dos *datasets* gerados nos testes iniciais deste trabalho. Tabela baseada em (Kerr et al., 2002).

Um método bastante simples para avaliar um modelo é dividir o dataset em dois grandes conjuntos (um para a construção do modelo - treino, e outro para testar o modelo - teste). Então se faz pelo menos uma medição estatística (por exemplo, acurácia) de cada instância do *dataset* de teste. Assim, pode-se utilizar o mesmo procedimento com outro modelo e realizar comparações. Mas esta técnica só é aceitável quando se tem um conjunto de dados grande o suficiente para realizar este tipo de repartição (Kotsiantis, 2007). Quando não se tem um dataset de tamanho avantajado, recorre-se a outros tipos

de métodos para verificar a generalização do modelo, tais como, *randomized-holdout* e *stratified-k-fold-cross-validation* (Witten e Frank, 2005).

No método *randomized-holdout* uma parte do dataset é escolhida randomicamente para treino e o restante é utilizado para teste, normalmente considerando a proporção de dois terço (treino) e o restante para teste. Este processo então é repetido diversas vezes e é computada a média das medidas estatísticas de cada repetição. Em cada repetição deve-se randomizar o dataset de maneira distinta (Witten e Frank, 2005).

No método *stratified-k-fold-cross-validation* é escolhido um número fixo k de *folds* (compartimentos ou pastas). O *dataset* é dividido randomicamente em k pastas iguais (ou aproximadamente), nas quais cada classe é representada aproximadamente na mesma proporção (estratificação - preserva a representatividade dos dados). Um compartimento é então utilizado para teste e, os $k-1$ restante são utilizados para treino, como pode-se observar na Figura 2.4. Daí computa-se uma medida estatística. Então o procedimento é executado k vezes. Por fim têm-se uma média das medidas estatísticas das k repetições. Assim como no *randomized-holdout* é recomendável que se repita o *stratified-k-fold-cross-validation* diversas vezes, com randomização diferente para cada repetição. Este método é um dos mais utilizados quando se tem uma base de dados pequena (Witten e Frank, 2005).

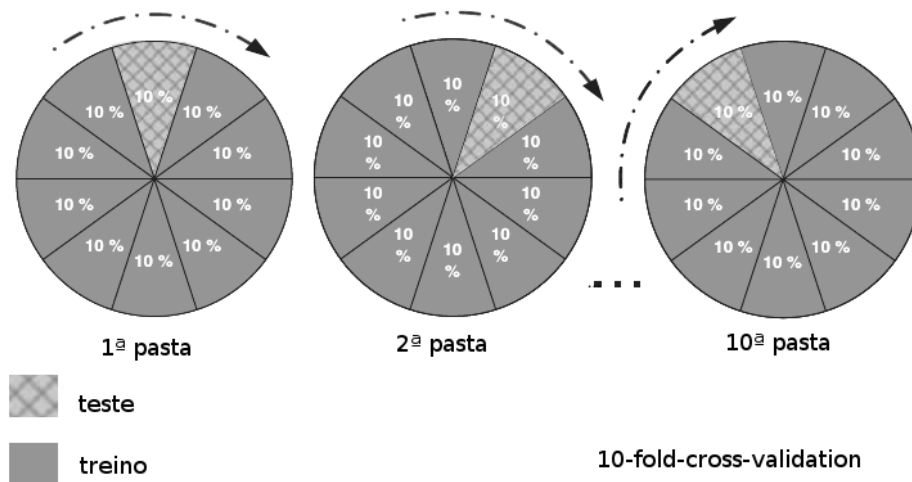


Figura 2.4: k -fold-cross-validation. Adaptado de (Olson e Delen, 2008).

Normalmente para o valor de k , no método descrito anteriormente, utiliza-se $k=10$. Este número de k tornou-se padrão após uma quantidade enorme de testes em datasets utilizando diferentes técnicas de classificação, os quais demonstraram $k=10$ como sendo aproximadamente a melhor escolha. A mesma quantidade (dez) têm sido escolhida como padrão para o número de repetições do *stratified-k-fold-cross-validation* (Witten e Frank, 2005).

No próximo capítulo (3) será apresentada a proposta deste trabalho para detectar efeitos colaterais indesejáveis entre dispositivos no cenário descrito na Seção 3.2.

Este capítulo descreve a proposta para detecção de efeitos colaterais indesejáveis no cenário “Levar compras”.

PROPOSTA

Diante da visão da IoT propõe-se detectar efeitos colaterais indesejáveis de uma maneira inteligente em um HNS, mais especificamente no cenário “Levar compras” (Seção 3.2). O cenário “Levar compras” foi desenvolvido com o intuito de criar um ambiente no HNS que tivesse efeitos colaterais indesejáveis, no qual os dispositivos são disponibilizados como serviços Web RESTFul, forma de disponibilização dos dispositivos que vêm sendo muito adotada (Mineraud et al., 2016). Como os dispositivos são disponibilizados como serviços Web, problemas relacionados a tal ambiente continuam a existir, já que ainda não foram sanados no ambiente dos serviços Web. Um destes é efeito colateral indesejável (Xu et al., 2010; Xu et al., 2011; Apel et al., 2014). Assim sendo, faz-se importante estudar metodologias possíveis para detectar e/ou resolver este problema, pois não se quer que uma composição de serviços (neste caso, dispositivos oferecidos como serviços) tenham um comportamento não esperado que leve a estados inconsistentes, dados imprecisos ou instabilidade.

Os efeitos colaterais indesejáveis que se quer verificar e detectar neste cenário são causados por violação de hipótese entre o HS e o elevador e, entre o elevador e a garra. Para verificar se acontece tais efeitos o cenário “Levar compras” é executado 59 vezes a fim de se observar e colher dados (Seção 3.3.1), os quais serão utilizados para se construir um modelo de *ensemble* final satisfatório a ser implantado no FIM do HNS. O *ensemble* utilizado será o DECORATE (Melville e Mooney, 2004), o qual utiliza o conhecimento de diversos classificadores para chegar em uma decisão final mais confiável. O método avaliativo utilizado para validar o modelo será o *stratified-k-fold-cross-validation*. Caso se chegue em um modelo final satisfatório, este será implantado no FIM do HNS.

3.1 ARQUITETURA

A Figura 3.1 demonstra a arquitetura do HNS e é descrita logo a seguir nas Seções 3.1.1, 3.1.2, 3.1.3, 3.1.4, 3.1.5, 3.1.6. Os serviços Web RESTFul que fazem parte do HNS são

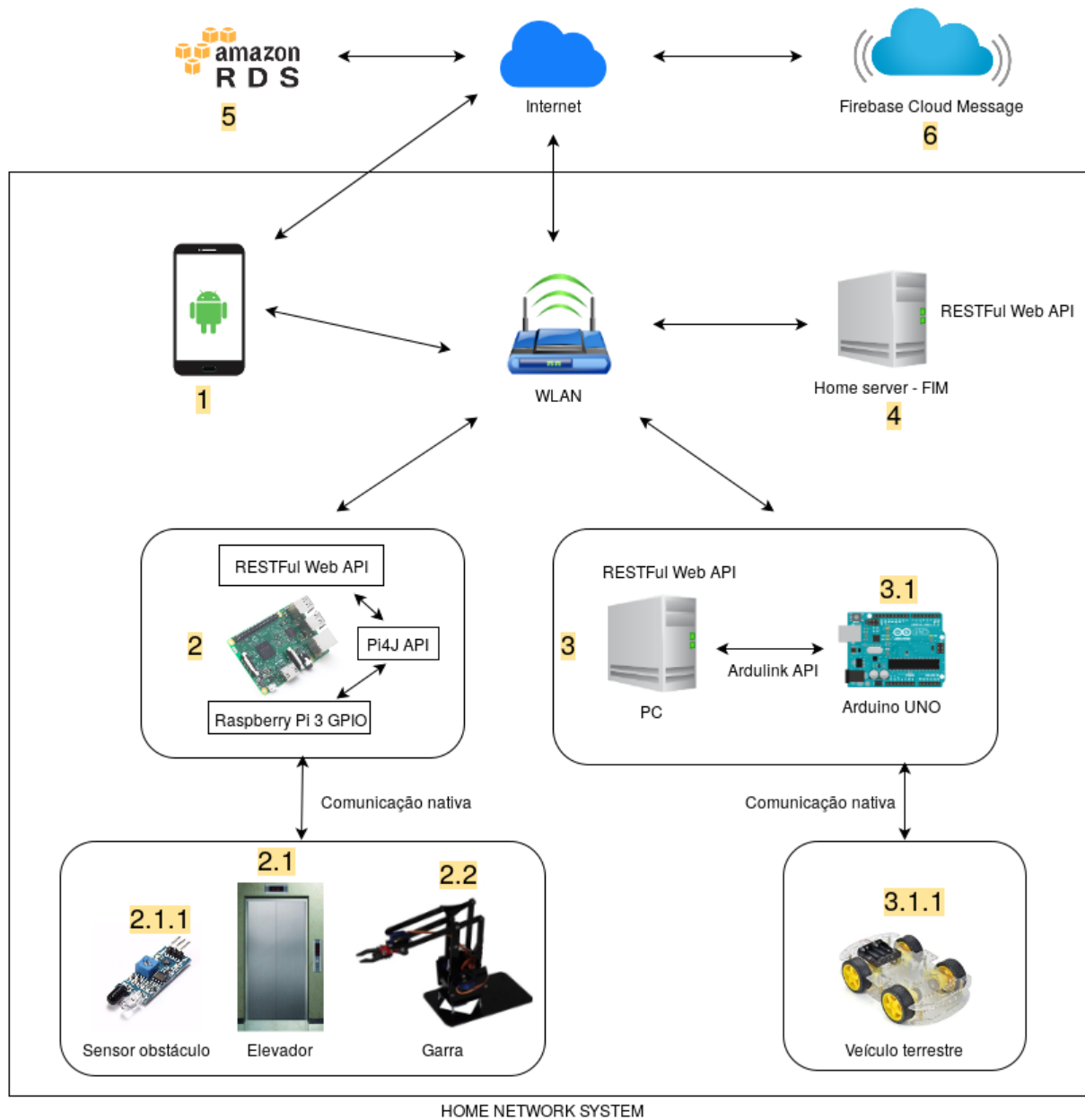


Figura 3.1: Modelo do HNS utilizado neste trabalho.

todos implementados utilizando a API Java JAX-RS, a qual prover o meio de utilizar os princípios REST (Heffelfinger, 2014). Estes serviços são hospedados no *container Glassfish 4.1*, uma aplicação servidora de código aberto que implementa as tecnologias Java EE (com completa referência). Glassfish permite que desenvolvedores desenvolvam e coloquem em produção aplicações Java EE, ou seja, ele prover as bibliotecas necessárias para desenvolver e colocar em produção aplicações Java seguindo as especificações Java EE (Heffelfinger, 2014). Os serviços trocam mensagens no formato JSON sobre o protocolo HTTP. O código fonte da implementação do HNS pode ser encontrado no repositório

heronsanches/final-paper¹.

3.1.1 Smartphone Android (1)

O *smartphone* com Android² possui um programa que solicita ao Home Server (4) que leve as compras, através da funcionalidade “Levar Compras”.

3.1.2 Raspberry Pi (2), elevador (2.1), garra (2.2) e sensor (2.1.1)

O Raspberry Pi (2) possui dois serviços, o do elevador (2.1) e o da garra (2.2), os quais são disponibilizados como serviços Web RESTFul independentes. O sensor de obstáculos é um componente somente do elevador, este não é disponibilizado como funcionalidade do serviço elevador. Uma funcionalidade não disponível ao usuário final (humano ou outra máquina) fica monitorando o estado do sensor (com obstáculo ou sem obstáculo), e qualquer mudança de estado é sempre informada ao elevador. A comunicação com os GPIOs do Raspberry Pi é realizada através da API Pi4J³. Os GPIOs do Raspberry Pi são que fazem a comunicação nativa com os dispositivos físicos (sensor - elevador, e servo motores - garra), é através dos GPIOs que consegue-se controlar e sentir os dispositivos.

3.1.3 PC (3) e Arduino UNO (3.1)

O PC (3) é utilizado para disponibilizar o veículo terrestre (3.1.1) como serviço Web RESTFul. O veículo terrestre não existe fisicamente, este é representado pelo Arduino via *software*, o qual simula duas ações do veículo, “ir para proximidades do elevador” e “cheguei nas proximidades do elevador”. A API Java Ardulink⁴ é utilizada para realizar a comunicação do PC com o Arduino através de sua porta serial, ela é utilizada para controlar e sentir o dispositivo, mais especificamente para solicitar que o veículo vá para as proximidades e escute quando o veículo chegou.

3.1.4 Home server - FIM (4)

O Home server - FIM (4) é disponibilizado como serviço Web RESTFul e oferece ao usuário da casa a funcionalidade “Levar compras”. O HS atua como FIM, é nele que será implementado o DECORATE (caso seja validado).

3.1.5 Firebase Cloud Message (6)

O Firebase Cloud Message⁵ (6) é utilizado pelo HS para enviar mensagens ao usuário da casa (1).

¹<https://github.com/heronsanches/final-paper>

²<https://www.android.com/intl/pt-BR.br/>

³<http://pi4j.com/>

⁴<http://www.ardulink.org/>

⁵<https://firebase.google.com/docs/cloud-messaging/?hl=pt-br>

3.1.6 Amazon RDS

Amazon RDS⁶ é um banco de dados relacional na nuvem, neste caso é utilizado o PostgreSQL⁷ com a finalidade de armazenar as cestas e objetos utilizadas para executar o cenário “Levar compras” (Seção 3.2.1) e, para armazenar a identificação do *smartphone* (1), a qual é utilizada para enviar mensagens ao usuário (1).

3.2 CENÁRIO “LEVAR COMPRAS”

O cenário “Levar compras” refere-se a incorporação de efeitos colaterais indesejáveis no HS. Esta funcionalidade é provida através da composição dos serviços do elevador, veículo e garra. O cenário em questão pode ser visualizado na Figura 3.2 e é descrito a seguir.

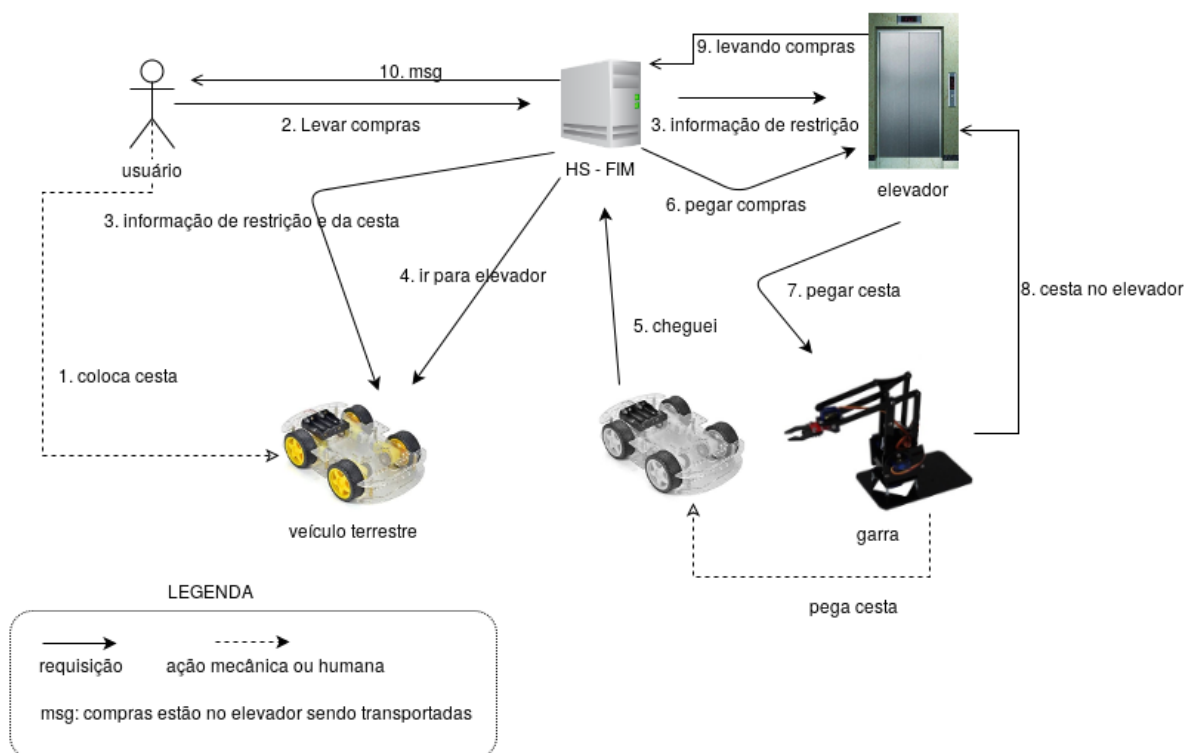


Figura 3.2: Cenário Levar compras.

1. Um usuário da casa (morador ou visitante), organiza as compras numa cesta (que tem formato expansível) e a coloca sobre o veículo terrestre.
2. O usuário então aciona o serviço “Levar compras” (disponibilizado pelo HS) através do seu *smartphone*. O serviço então começa sua execução.
3. O FIM solicita informações de restrições do elevador e veículo terrestre, assim como quais objetos compõem a cesta. Este então verifica se a cesta colocada sobre o carro

⁶<https://aws.amazon.com/pt/rds/postgresql/>

⁷<https://www.postgresql.org/>

quebrou alguma restrição do mesmo e se a massa total dos objetos excedem a carga máxima do elevador.

4. Se nenhuma destas restrições foram quebradas o FIM solicita ao veículo terrestre que leve a as compras até as proximidades do elevador.
5. Assim que o veículo chega no seu destino (proximidades do elevador), este envia uma mensagem ao FIM informando de sua chegada.
6. O FIM então requisita o elevador para levar as compras.
7. O elevador agora solicita a garra mecânica que pegue o cesto e coloque-o dentro do elevador.
8. Quando a garra completa a execução do seu serviço, esta avisa ao elevador.
9. Caso não haja nenhuma restrição, como por exemplo, alguma coisa que possa impedir a porta do elevador de fechar, então o elevador inicia o transporte das compras ao seu destino final e manda uma notificação ao HS.
10. HS Captura token referente ao *smartphone* com Android do usuário da casa. Por fim, o HS solicita ao serviço de notificação *Firebase Cloud Message* que envie uma notificação para o usuário da casa, que a recebe no seu *smartphone*. Assim completa-se a execução do serviço “Levar compras”.

No cenário descrito da Figura 3.2 o elevador não contém motores nem algum outro tipo de sistema elétrico ou mecânico, este é representado por uma caixa de papelão e um sensor de obstáculo acoplado. O veículo terrestre é representado por um programa carregado no Arduino UNO⁸, o qual escuta através de sua porta serial mensagens vindas do PC (Figura 3.2) e as processa simulando as funcionalidades “Ir para elevador” e “cheguei”. Pela mesma porta serial o veículo terrestre encaminha suas mensagens ao PC. Os dispositivos veículo, elevador e garra podem ser visualizados na Figura 3.3.

3.2.1 Execução do cenário Levar compras

De acordo com a Figura 3.2 o primeiro passo para inicializar a execução do cenário é colocar a cesta no veículo terrestre. Apesar do veículo ser simulado através de um software embarcado no Arduino, este necessita das informações dos objetos que se deseja transportar. Sendo assim, as informações da cesta são passadas quando o usuário executa o passo 2 (Levar compras). O usuário digita o “id da cesta” (do banco de dados) que se deseja transportar e clica no botão (representado por uma caixa) para inicializar o serviço, como pode ser visualizado na Figura 3.4. Todas as cestas são compostas com miniaturas de objetos que normalmente são encontrados em supermercados, exemplos de cestas podem ser visualizados na Figura 3.5. Todos os objetos e cestas estão cadastradas no banco de dados na nuvem (amazon RDS), o qual foi apresentado na Figura 3.1. A lista

⁸<https://www.arduino.cc/en/Main/ArduinoBoardUno>

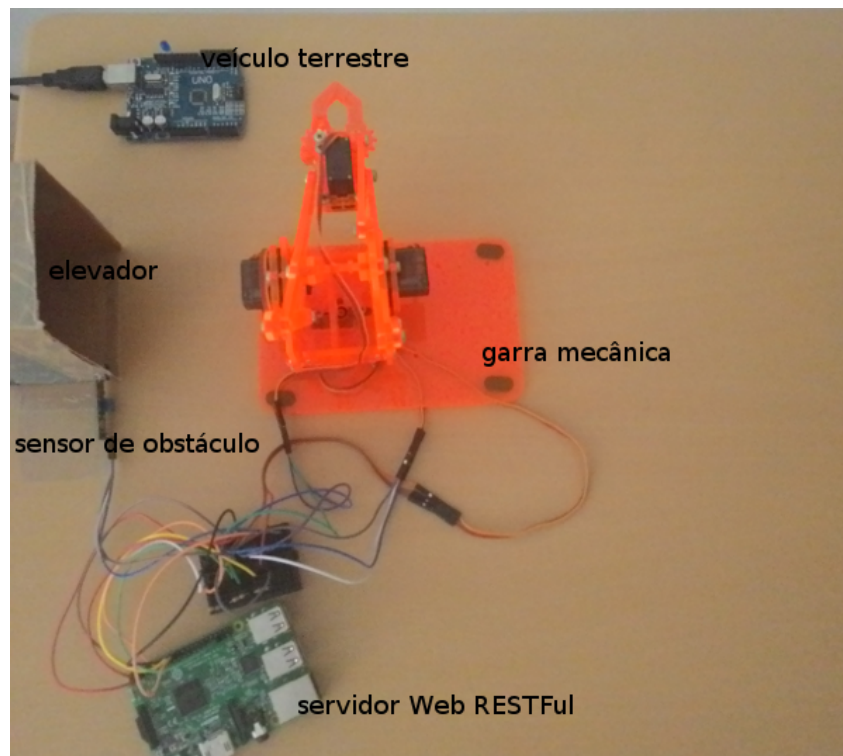


Figura 3.3: Dispositivos reais do cenário “Levar compras”.



Figura 3.4: Tela da aplicação Android do usuário da casa.

de objetos que podem fazer parte de uma cesta é demonstrada na Figura 3.6, observe que para cada objeto é cadastrado suas características de massa, largura, comprimento, altura e diâmetro.

Os próximos passos (3, 4, 5, e 6) do cenário já têm a informação necessária e não necessita de nenhuma intervenção humana. Isto já não acontece com o passo 7 (pegar cesta), pois este necessita de intervenção humana para a garra poder pegar a cesta e continuar a realizar o seu serviço, é preciso posicionar a cesta no lugar correto para a



Figura 3.5: Exemplos de cestas do cenário “Levar compras”.

largura	altura	comprimento	massa	diâmetro	nome
0	0.031	0	0.003	0.009	Refrigerante nº 4 - Soda
0.01	0.014	0.03	0.001	0	Caixa Pasta de dente - vermelha.
0.004	0.02	0.014	0.001	0	Batata
0.004	0.02	0.014	0.001	0	Café
0	0.015	0	0.003	0.008	Café Classic
0.005	0.02	0.011	0.001	0	Cerveja em lata III
0	0.028	0	0.001	0.007	Champanhe - Azul
0.004	0.02	0.014	0.001	0	Farinha de Trigo
0	0.03	0	0.001	0.008	Garrafa de Rum
0.012	0.016	0.004	0.004	0	Goiabada
0	0.03	0	0.001	0.008	Garrafa de Vodka Special
0.027	0.014	0.018	0.001	0	Pote de sorvete
0	0.032	0	0.002	0.007	Vinho branco
0.012	0.03	0.006	0.002	0	Limpa vidros
0.021	0.015	0.007	0.002	0	Sabonete - II
0.015	0.024	0.008	0.001	0	Sabão em pó - II
0.019	0.03	0.039	0.018	0	Bau simples
0.046	0.061	0.017	0.006	0	Prateleira de acrílico pequena
0.016	0.025	0.004	0.001	0	Farinha de trigo
0	0.03	0	0.003	0.022	Barril em madeira
0.013	0.017	0.0063	0.002	0	Fermento em pó
0.077	0.0197	0.038	0.001	0	Colchão p/ berço azul
0.01	0.0432	0	0.001	0	Jogo de esfregões
0.0181	0.0258	0.0017	0.001	0	Pão de Queijo
0.022	0.0412	0.0135	0.001	0	Rastelo

Figura 3.6: Lista dos objetos que podem ser utilizados para compor uma cesta.

garra prender a cesta, como pode ser visto na Figura 3.7. Então, após a garra executar seu serviço esta manda uma mensagem para o elevador e, caso não tenha nada obstruindo a porta do elevador este inicia o transporte da cesta até o local desejado e manda uma mensagem para o FIM. Daí em diante a execução do cenário continua da mesma forma com já explanado.

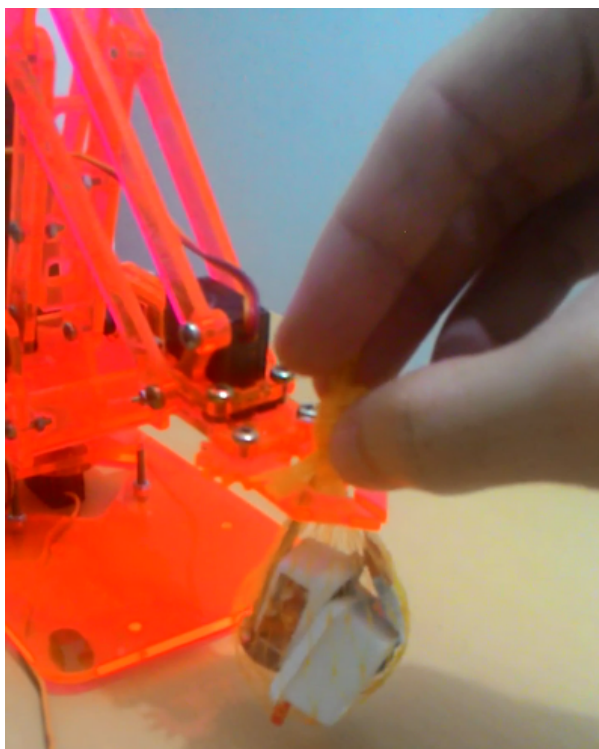


Figura 3.7: Cesta posicionada no local adequado por um humano, para que a garra possa prender a cesta adequadamente.

3.3 MÉTODO DE DETECÇÃO

3.3.1 Geração da massa de dados

A geração da massa de dados para o desenvolvimento do método proposto de detecção ocorreu de forma simulada, já que os dispositivos não são todos reais. Os dados gerados foram obtidos através da execução do cenário “Levar compras”. Para a execução deste cenário foram confeccionadas e cadastradas no banco de dados 59 cestas distintas. Antes de registrar as cestas no banco sempre era observado se a cesta cabia dentro do elevador quando colocadas de forma manual, e se também não ultrapassava os limites da porta, o limite de detecção do sensor de obstáculos. Então a cesta só era registrada caso coubesse no elevador e não estivesse em área que pudesse obstruir a porta. Diante das observações da execução do cenário, percebeu-se que as informações das cestas poderiam ser um indicativo ao problema (efeito colateral indesejável) provocado pelo comportamento da garra. Como todas as cestas estavam previamente cadastradas no banco de dados, a estas foram atribuídos valores de “é efeito colateral” ou “não é efeito colateral”, à medida que se observava cada execução do cenário. Desta forma, para construção do modelo (Seção 4.1) que classifique o problema em “é efeito colateral” ou “não é efeito colateral” utilizou-se as informações das cestas. Vale salientar que a informação da cesta é uma resposta do veículo terrestre referente a requisição “3” da Figura 3.2, ou seja, é uma das informações que os dispositivos e/ou serviços trocam entre si durante a execução do cenário.

O cenário foi executado 59 vezes, uma para cada cesta, e pode-se observar que em algumas vezes o comportamento da garra provocava efeitos colaterais indesejáveis (Seção 2.2) à funcionalidade “Levar compras”, pois para todos esses casos se esperava sempre como resultado o elevador levar as compras sem problemas, já que era sabido que as cestas cabiam no elevador e, nenhuma delas feria as restrições de massa do carro nem do elevador.

O efeito colateral indesejável acontece em dois momentos da execução do cenário e em ambos pela mesma causa (violação de hipótese), verificando assim a proposta apresentada para este cenário.

1. O HS supôs que o elevador poderia levar as compras sem problemas, já que se era sabido que as compras cabiam no elevador e não feria nenhuma restrição de massa, mas alguma vez o elevador não conseguia levar as compras, pois estas ficavam impedindo o elevador de fechar a porta.
2. O elevador supôs que o serviço da garra seria apto a colocar as compras dentro do seu compartimento sem problemas, mas algumas vezes os objetos transportados sofriam movimentação na cesta (uma sacola flexível) durante a locomoção e, em especial, no momento em que a garra soltava a cesta no elevador, esta as vezes ganhava novas dimensões ou sofria locomoção para a linha de detecção do sensor de obstáculo da porta do elevador, devido a natureza dos objetos e da forma como os objetos eram soltos pela garra. Por esse motivo utilizou-se a cesta como informação para construção do modelo do *ensemble* DECORATE.

3.3.2 Seleção dos atributos

Os dados coletados na etapa de execução do cenário “Levar compras”, 59 cestas, ainda estão muito brutos. Uma cesta pode conter um objeto, como também pode conter muitos. Supomos que a cesta tenha 10 objetos, então o total de características da cesta seria “ $10 \times 5 = 50$ ”, pois cada objeto possui cinco características. Utilizar as características dos objetos desta forma terá um custo computacional muito elevado. Para diminuir a quantidade de atributos e melhor qualificá-los gerou-se características a partir de todos os atributos dos objetos da cesta: média das alturas, dos comprimentos, das larguras, dos diâmetros e das massas, assim como o total de cada uma dessas medidas, e o total de itens. De início então têm-se a seleção de 11 parâmetros.

Com a seleção prévia desses atributos construiu-se um arquivo “arff”, que é o tipo de arquivo padrão utilizado como *dataset* pelo Weka (Hall et al., 2009). Weka é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados, este possui ferramentas para pré-processamento de dados, classificação, agrupamento, dentre outros. O *dataset* construído é composto de 59 instâncias (referente as 59 cestas), cada uma com 11 atributos mais um que identifica a qual classe pertence (“é efeito colateral” ou “não é efeito colateral”). Então têm-se 11 atributos numéricos e um nominal (a classe a qual a instância pertence).

O próximo passo utilizado nesse processo de seleção de atributos foi o de visualizar estes novos parâmetros par a par (Figura 3.8), e verificar quais pares parecem melhor

separar as classes “é efeito colateral” ou “não é efeito colateral”. Cada quadro nessa Figura representa um par de atributos, por exemplo, “somaDiâmetro x somaLargura”, o qual mostra a separação das classes, cada uma representada por uma cor diferente. A cor azul representa “não é efeito colateral” e a vermelha “é efeito colateral”.

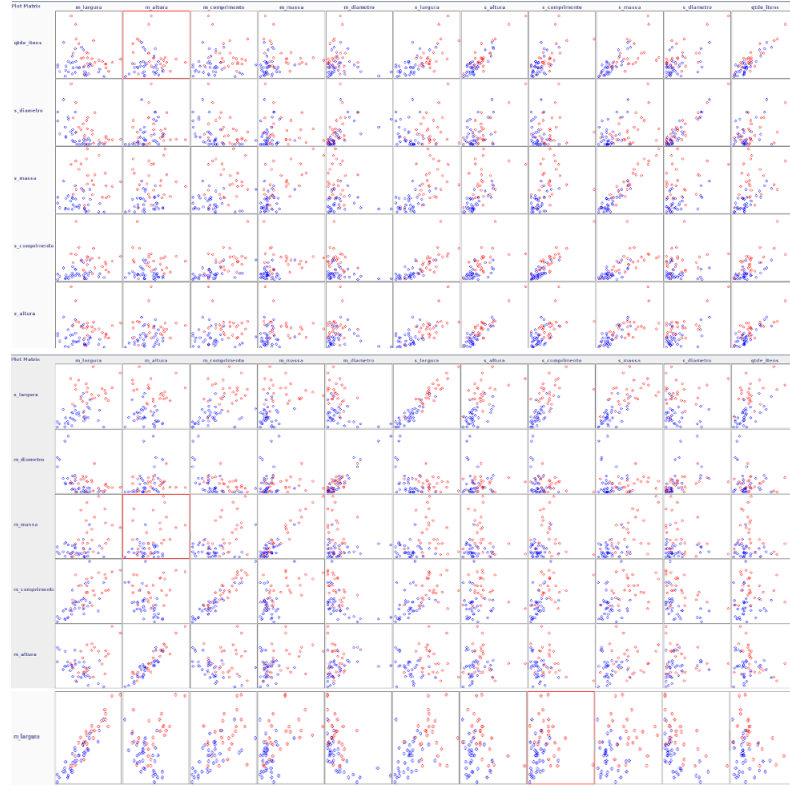


Figura 3.8: Plotagem do espaço de *features* par a par do dataset de 59 instâncias.

Dentre essas plotagens foram selecionadas três (Figura 3.9), as quais aparentam melhor separar as classes, são respectivamente os pares ordenados (somaLargura, somaMassa) - “sl x sm”, (somaDiâmetro, somaLargura) - “sd x sl” e, (médiaAltura, somaLargura) - “ma x sl”.

Então, nesta fase de seleção dos atributos, estes foram os pares selecionados para construção e validação do modelo (Seção 4.1). Observe que o espaço de *features* foi reduzido a duas características.

3.3.3 Meta-classificador DECORATE

Os parâmetros do DECORATE escolhidos neste trabalho foram semelhantes a um dos testes avaliativos realizados em (Melville e Mooney, 2004), os quais comparam *DECORATE* com outros métodos classificatórios, em diferentes tamanhos de dataset e diferentes parâmetros do DECORATE. Para as avaliações neste trabalho o número de iterações é setado em 50, o número máximo de classificadores participantes no *ensemble* é setado em 14, o classificador base utilizado é o J48 e, a quantidade de exemplos artificiais gerados

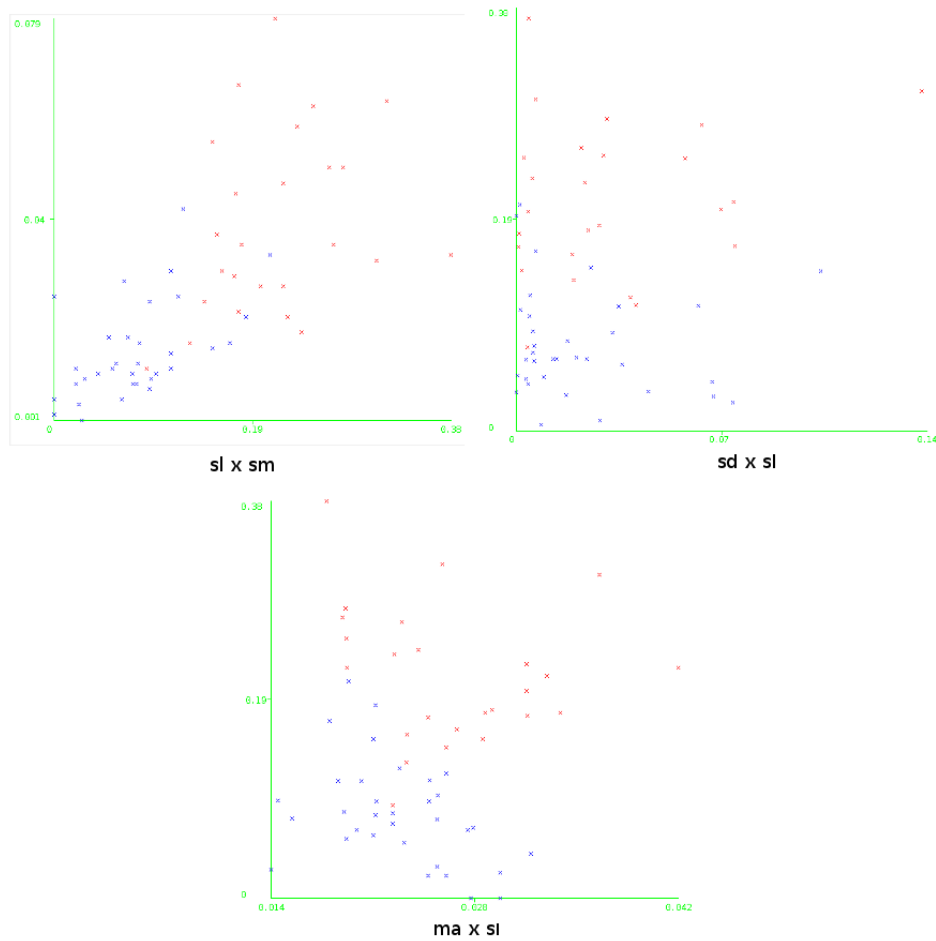


Figura 3.9: Separação das classes entre os atributos (somaLargura, somaMassa) - “sl x sm”, (somaDiâmetro, somaLargura) - “sd x sl” e, (médiaAltura, somaLargura).

em cada iteração é setado para 100% (valores setados de 50% a 100% não faz variar muito o resultado (Melville e Mooney, 2004)) do tamanho do dataset de treino.

3.4 IMPLANTAÇÃO DO MÉTODO PROPOSTO

O classificador atua exatamente no momento em que o serviço veículo responde a requisição do FIM na etapa 3 da Figura 3.10. Nesta etapa o veículo primeiramente captura o “id_cesta” (banco de dados) passado a ele pelo FIM. Então o veículo utiliza uma funcionalidade do serviço FIM que captura a cesta correspondente ao “id_cesta”, transforma-a em um arquivo “arff” contendo somente uma instância com os parâmetros correspondentes a “somaLargura, mediaAltura, classe”, sendo a classe com valor não rotulado, ou seja, não sabe-se a qual classe pertence (“é efeito colateral” ou “não é efeito colateral”). O FIM então salva este arquivo e retorna o caminho do arquivo como resposta para o dispositivo veículo. O veículo então responde a requisição 3, sendo um dos dados o caminho do arquivo “arff” referente a cesta. Neste momento o FIM utiliza o *ensemble*

DECORATE (implantado como uma funcionalidade interna) para predizer se “é efeito colateral” ou “não é efeito colateral” passando como parâmetro ao DECORATE o caminho do arquivo “arff”. Caso o classificador classifique como “é efeito colateral” o FIM interrompe o serviço e manda uma mensagem para o usuário informando que ocorreu “efeito colateral indesejável”. Caso contrário o fluxo do cenário (Figura 3.2) continua e no final da execução o usuário recebe uma mensagem de sucesso.

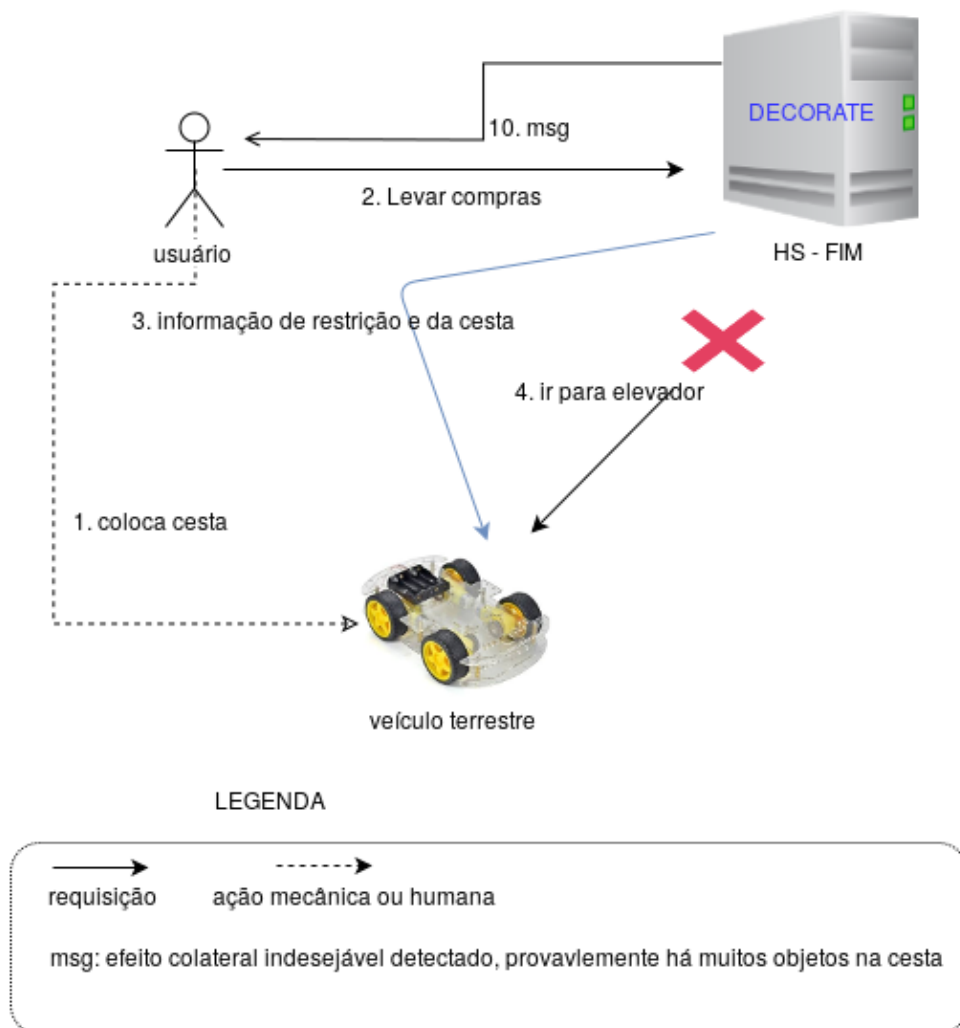


Figura 3.10: Implantação do método proposto.

Procedimentos realizados para validar a proposta de pesquisa deste TCC.

VALIDAÇÃO

4.1 EXPERIMENTOS

Diante dos pares de atributos selecionados na etapa (seção 3.3.2), utilizou-se o método *stratified-10-fold-cross-validation* (repetido dez vezes) juntamente com o *ensemble DECO-RATE* a fim de se construir e verificar se é possível generalizar pelo menos um modelo.

Em cada verificação construi-se um modelo com diferentes frações do dataset, variando de 16 instâncias a 59 instâncias. Os resultados das avaliações podem ser vistos na Figura 4.1 e será discutido logo a seguir na seção 4.2.1.

4.2 RESULTADOS E DISCUSSÃO

4.2.1 Modelos

A Figura 4.1 demonstra os resultados obtidos através dos experimentos realizados (seção 4.1). Nesta, percebe-se um domínio quase absoluto da curva pontuada de aprendizado verde (médiaAltura, somaLargura - “ma x sl”). Por isso motivo essa será a curva analisada para escolher um modelo o mais adequado possível para implantar no FIM.

Como o principal motivo da implantação do classificador no FIM é detectar efeitos colaterais indesejáveis, então pretende-se que o modelo tenha um alto índice de *recall*, pois não se deseja que se tenha muitos falso negativos, ou seja, muitos casos classificados como não sendo efeito colateral, mas que na verdade era. Visualizando a curva percebe-se que esta têm um crescimento acentuado e depois torna-se estável a partir do dataset com 30 instâncias. Logo será realizada uma análise a partir deste ponto. Os maiores valores de *recall* partindo deste ponto são aqueles quando o *dataset* tem 40 ou mais instâncias, para todos estes pontos considera que se obteve um bom grau de generalização. Estes valores podem ser visualizados na Tabela 4.1.

Finalizando a análise decidiu-se que o melhor ponto para poder se construir um classificador é o com $n^oi = 45$, pois além de se ter um alto grau de *recall* e *accuracy* têm-se também *precision* quase de 100%. Então, para se construir um classificador final com

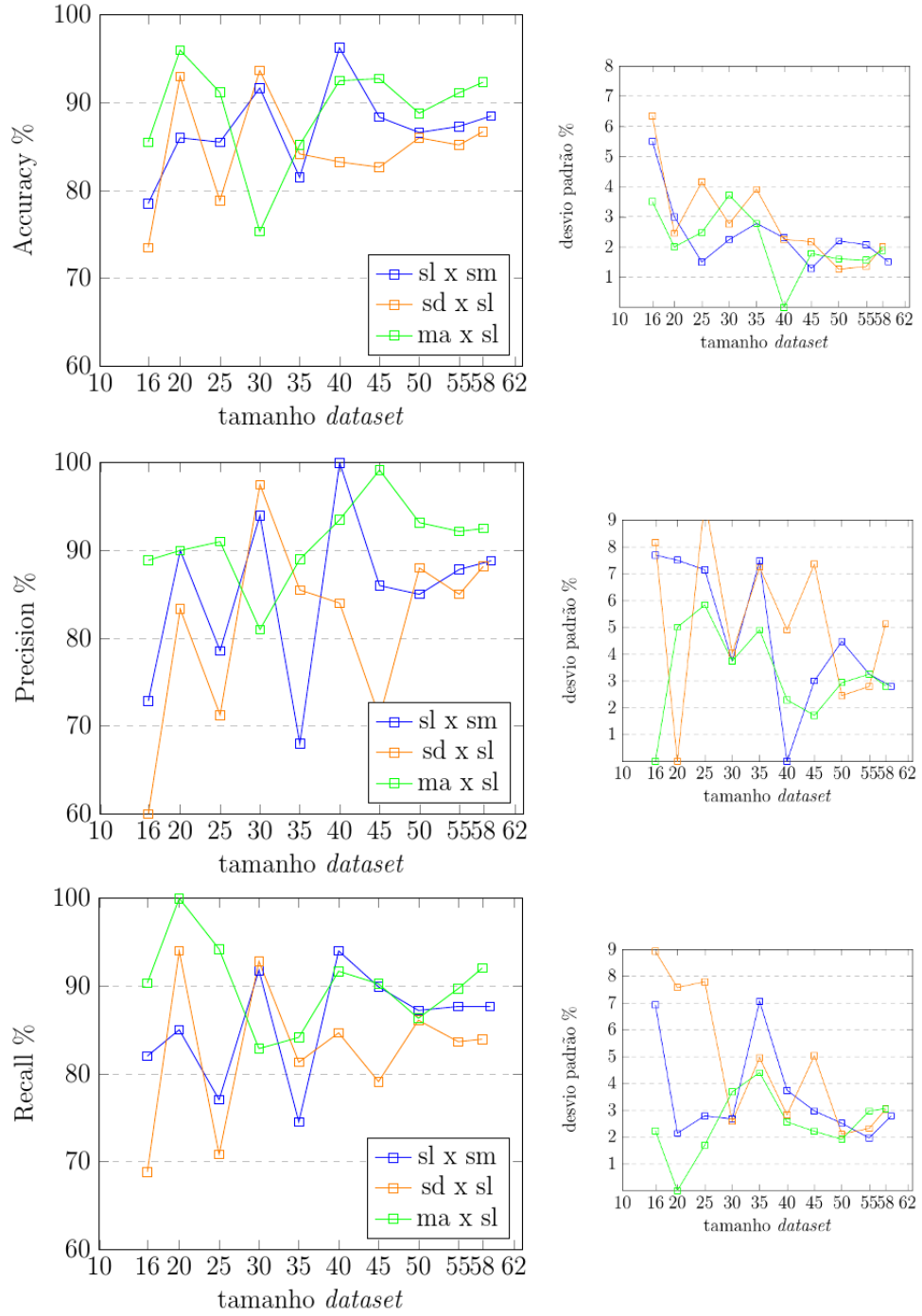


Figura 4.1: *Accuracy*, *Precision* e *Recall* obtidas utilizando stratified-10-fold-crossvalidation (repetido 10 vezes) e DECORATE sobre diferentes frações do *dataset* com os pares de parâmetros selecionados na seção 3.3.2.

base neste modelo escolheu-se randomicamente 45 instâncias do *dataset* de 59 instâncias e aplicou-se *stratified-10-fold-cross-validation* (repetido dez vezes). Este procedimento foi repetido diversas vezes (Figura 4.2) até que se obtivesse um modelo que satisfizesse a

n^oi	40	45	50	55	58
A, DP	92.5, 0	92.75, 1.78	88.8, 1.6	91.13, 1.56	92.33, 1.89
P, DP	93.5, 2.29	99.17, 1.71	93.17, 2.93	92.17, 3.25	92.5, 2.81
R, DP	91.69, 2.56	90.25, 2.21	86.33, 1.91	89.71, 2.97	92.05, 3.06

Tabela 4.1: Valores das curvas “ma x sl” da Figura 4.1 a partir do *dataset* com 40 instâncias. A (*Accuracy*), P (*Precision*), R (*Recall*), DP (Desvio Padrão), i (instâncias). Os valores A, P, R e DV estão em %.

expressão 4.1.

$$(accuracy \geq (92.75 - 1.78)) \wedge (precision \geq (99.17 - 1.71)) \wedge (recall \geq 90.25) \quad (4.1)$$

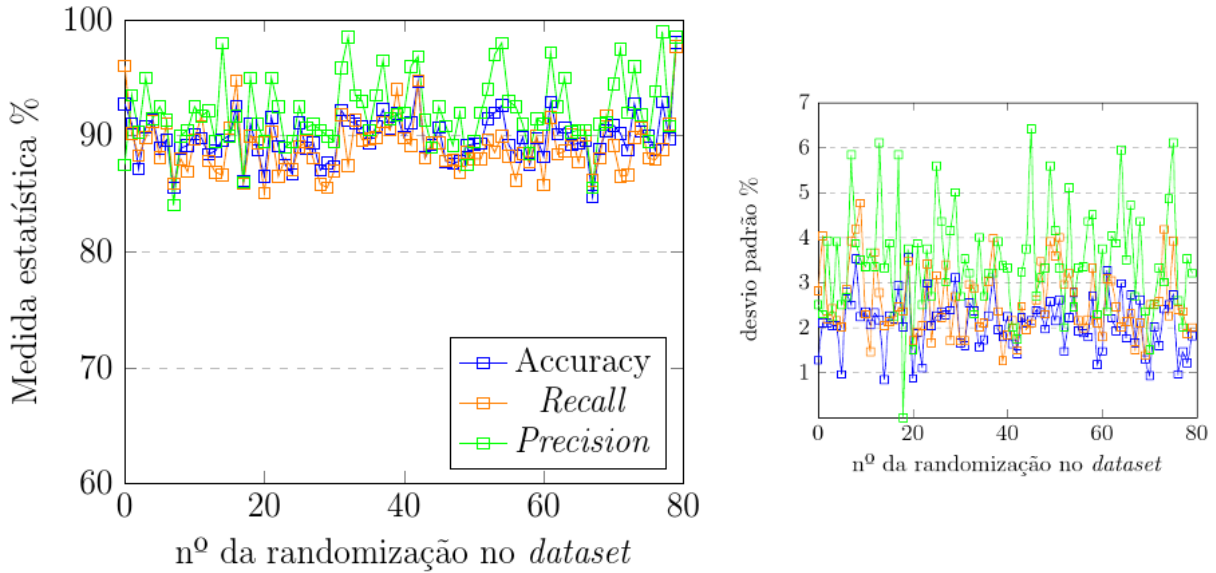


Figura 4.2: *stratified-10-fold-cross-validation* (repetido dez vezes). 45 instâncias foram escolhidas randomicamente em cima do *dataset* de 59 instâncias. Procedimento repetido até que satisfizesse a expressão 4.1

O modelo que satisfaz essa expressão obteve resultados que podem ser vistos na Tabela 4.2 e este será o modelo final de classificador que será implantado no FIM.

		DP %
Accuracy %	98.05	1.81
Precision %	98.5	3.2
Recall %	97.67	2

Tabela 4.2: Valores da validação do modelo final. DP (Desvio Padrão).

4.2.2 Implantação

O modelo final construído foi implantado no sistema de acordo com a proposta da seção 3.4. Na Figura 4.3 se tem a visão real do fluxo de execução em relação ao usuário do HNS quanto a funcionalidade “Levar compras”. A imagem “a)” representa a execução quando não há efeitos colaterais indesejáveis. Enquanto que a imagem “b)” representa a execução quando se tem efeito colateral indesejável.

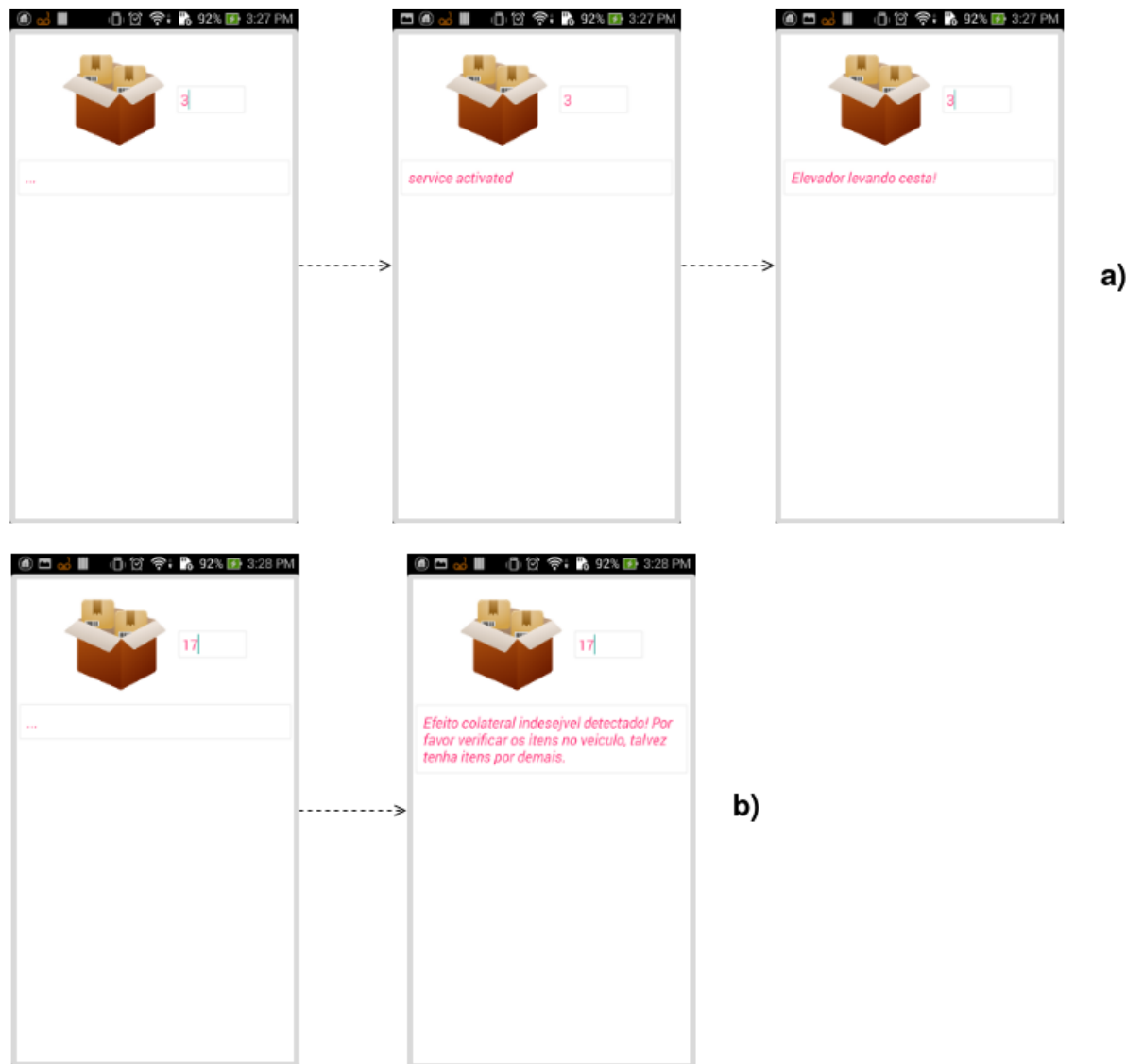


Figura 4.3: Modelo classificatório atuando no cenário “Levar compras” do FIM. a) Fluxo da aplicação sem efeito colateral indesejável. b) Fluxo da aplicação com efeito colateral indesejável.

Descreve de forma resumida os trabalhos relacionados ao desta monografia, assim como em que difere destes.

TRABALHOS RELACIONADOS

(Maternaghan e Turner, 2013) propõem uma abordagem *offline* (na fase de construção do sistema) para detecção de efeitos colaterais indesejáveis causados por conflitos de políticas. Nesta abordagem é verificado as ações de cada política de forma isolada e os pares de políticas que causam conflitos.

Em (Wilson et al., 2008) é apresentado um modelo de gerenciamento de Unidade de Funcionalidade de Software (UFS) ou *feature* em um *Home Automation System* para detecção de interação de características. Caso um efeito colateral indesejado ocorra, este deve ser evitado, mas se ocorre um efeito colateral desejado, este deve ser permitido. Este modelo consiste em três camadas de cima para baixo, como segue.

- Camada superior (serviços que automatizam a casa) - utiliza um ou mais dispositivos, os quais estão localizados na camada intermediária (dispositivos);
- Camada de dispositivos – contem dois tipos de dispositivos (entrada (ex.: termômetro, somente monitora um aspecto do ambiente e retorna para os serviços) e saída (ex.: aquecedor, o qual altera o ambiente))
- Camada de ambiente – contém variáveis de ambiente, por exemplo: movimento do quarto, temperatura do quarto, luminosidade do quarto, umidade do quarto, fumaça do quarto.

O gerenciamento de UFS trabalha controlando o acesso à camada de ambiente utilizando de variáveis de bloqueio de acesso, especificando prioridades de serviços e além disso se utiliza de um banco de dados na nuvem para manter detalhes a respeito de cada dispositivo, com a finalidade de entender o efeito das ações de cada dispositivo no ambiente.

Os autores em (Nakamura et al., 2009) apresentam um método de detecção e resolução online para efeitos colaterais indesejáveis entre serviços integrados no *Home Network System*. Os autores definem dois tipos de efeitos colaterais indesejáveis, *appliance interaction* e *environment interaction*. *Appliance interaction* refere-se a uma situação onde dois

serviços compartilham um mesmo dispositivo de forma conflitante, enquanto que *environment interaction* ocorre quando dois serviços, os quais não necessariamente compartilham os mesmos dispositivos, mas conflitam-se por propriedades de ambientes, como, por exemplo, luminosidade. O método proposto utiliza mecanismos de suspender/resumir, prioridade e tempo de ativação dos serviços para detectar e resolver as interações. Em outro trabalho mais adiante (Nakamura et al., 2013), no qual Nakamura e Ikegami participaram, estes com Matsumoto propuseram uma extensão do trabalho anterior. Nesta, os autores definem um modelo (*environment impact model*) o qual define como cada dispositivo contribui para as variáveis de ambiente. Também é introduzido um *environment requirement* para definir o estado esperado de cada serviço. Então o conceito anterior de *environment interaction* é reformalizado introduzindo uma condição que uma certa quantidade de impactos acumulados viola os requerimentos dos serviços em questão. O autor realizou cinco casos de estudos os quais conseguiu detectar interações que não tinha conseguido com a definição anterior.

(Alfakeeh e Al-Bayatti, 2016) utilizam um método de detecção e resolução online o qual é um mecanismo de negociação entre os serviços ou as preferências do usuário da casa a fim de que seja possível estes serviços trabalharem em conjunto ao mesmo tempo. Este mecanismo de negociação é um *Agent-Based Negotiation System* (ABNS) de detecção e resolução de efeitos colaterais indesejáveis na casa inteligente.

A proposta apresentada neste TCC apresenta um método de detecção online, assim como em (Wilson et al., 2008; Nakamura et al., 2009; Nakamura et al., 2013; Alfakeeh e Al-Bayatti, 2016). Nos trabalhos citados, na arquitetura de HNS utilizada, os dispositivos são utilizados para prover serviços ao usuário da casa, e estes são acessados somente pelo HS através das APIs dos dispositivos. Os dispositivos nesta arquitetura não têm capacidade de se comunicar e oferecer serviços entre eles de forma independente. No trabalho desenvolvido neste TCC, os dispositivos são disponibilizados como serviços Web RESTful e estes têm capacidade de se comunicar e oferecer serviços de forma independente, representando desta forma um cenário da IoT. Além disto, o método de detecção online proposto no Capítulo 3 utiliza aprendizagem de máquina para detecção dos efeitos colaterais indesejáveis, mais precisamente se utilizou do *ensemble* DECORATE para prover tal capacidade.

Síntese da investigação e dos experimentos realizados nesta monografia.

CONCLUSÃO

O presente trabalho demonstrou através de um cenário (“Levar compras” - Seção 3.2) no HNS que quando dispositivos são compostos para prover uma única funcionalidade, o comportamento de pelo menos um destes pode provocar interação de características com efeitos colaterais indesejáveis, no caso deste cenário, provocados por violação de hipótese entre o HS e o elevador e, entre o elevador e a garra. Os dispositivos do cenário “Levar compras” são disponibilizados como serviços Web RESTFul, então estes têm capacidade de se comunicar, ser identificado unicamente e oferecer serviços de forma independente, representando desta forma um cenário da IoT.

Para o cenário em questão foi proposto detectar tais efeitos colaterais indesejáveis, de forma online, utilizando-se de um *Feature Interaction Manager* (FIM). Para que o FIM pudesse realizar a detecção dos efeitos colaterais indesejáveis foi implantado um modelo do *ensemble* (DECORATE) na funcionalidade “Levar compras”, este atua justamente no momento em que o FIM pede informações de restrição e conteúdo da cesta que está sobre o carro. O FIM então pega as informações do conteúdo da cesta, transforma-as em formato de dados que o DECORATE entenda (neste caso em específico, um arquivo arff) e realiza a classificação em “é efeito colateral” ou “não é efeito colateral”.

O modelo implantado no FIM foi concebido através de um processo classificatório (Figura 2.2), no qual as etapas podem ser representadas pelas seções (3.2, 3.3.2, 4.1). Na etapa de validação utilizou-se o método *stratified-10-fold-crossvalidation* repetido dez vezes diante diferentes porções do *dataset*, com o intuito de saber a partir de quantas instâncias o modelo iria generalizar. Ficou observado que o melhor resultado obtido foi quando o modelo construído utilizou-se de 45 instâncias do total de 59 instâncias do *dataset*. Então para construir o modelo final, o qual foi implantado no FIM, executou-se o *stratified-10-fold-crossvalidation* (repetido 10 vezes) por diversas vezes sobre 45 instâncias do *dataset* (escolhidas aleatoriamente com seeds diferentes) até que a expressão 4.1 fosse satisfeita, esta é baseada nos resultados obtidos da Tabela 4.1 com *dataset* de tamanho 45. O resultado da expressão satisfeita gerou um modelo classificatório final com *Accuracy* = 98.05%, *Precision* = 98.5% e *Recall* = 97.67%.

Assim, o FIM ficou apto a realizar detecção inteligente de efeitos colaterais indesejáveis para o cenário proposto, mostrando que é possível detectar efeitos colaterais indesejáveis entre dispositivos na visão da IoT.

Como trabalhos futuros, pretende-se:

- Realizar um estudo sobre algoritmos de seleção de atributos afim de não mais selecionar um vetor de atributos (que descreve o problema) de forma manual. Observe que os atributos selecionados de forma manual e visual (Seção 3.3.2) para o cenário proposto neste trabalho estavam em um espaço de 2 dimensões (Figura 3.8), algumas vezes um espaço de duas dimensões não é suficiente para descrever um problema, o mesmo acontece para um espaço 3D (o máximo que a visão humana consegue visualizar). Realizado este estudo, então pretende-se executar os experimentos realizados neste trabalho (Seção 4.1) em cima dos vetores de atributos selecionados pelos algoritmos estudados e realizar comparações dos resultados obtidos na Seção 4.2.1 com cada um dos algoritmos de seleção de atributos utilizados.
- Criar um cenário, com mais dispositivos, que seja o mais real possível e com pouca intervenção humana ou nenhuma e utilizar os algoritmos de seleção de atributos estudados e realizar as mesmas comparações.
- Estudar sobre aprendizado não supervisionado e prover metodologia para realizar classificação de efeitos colaterais indesejáveis no cenário deste trabalho e no outro que pretende-se criar citado anteriormente.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALFAKEEH, A. S.; AL-BAYATTI, A. H. Feature interactions detection and resolution in smart homes systems. *International Journal of Electronics and Electrical Engineering*, v. 4, n. 1, 2016.
- ALMEIDA, J. et al. An overview of formal methods tools and techniques. In: _____. *Rigorous Software Development: An Introduction to Program Verification*. London: Springer London, 2011. p. 15–44. ISBN 978-0-85729-018-2. Disponível em: http://dx.doi.org/10.1007/978-0-85729-018-2_2.
- APEL, S. et al. Feature interactions: the next generation (dagstuhl seminar 14281). *Dagstuhl Reports*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, v. 4, n. 7, 2014.
- ARORA, R.; SUMAN, S. Comparative analysis of classification algorithms on different datasets using weka. *International Journal of Computer Applications*, v. 54, n. 13, 2012.
- ASHTON, K. That 'internet of things' thing. *RFID Journal*, 2009.
- ATZORI, L.; IERA, A.; MORABITO, G. The internet of things: A survey. *Computer Networks*, v. 54, p. 2787– 2805, October 2010.
- BELQASMI, F. et al. Soap-based web services vs. restful web services for multimedia conferencing applications: A case study. *IEEE INTERNET COMPUTING*, 2012. Disponível em: <http://spectrum.library.concordia.ca/980040/1/PersonalCopy-SOAPvsREST.pdf>.
- BLAND, J.; ALTMAN, D. Measurement error. *BMJ*, v. 313, p. 744, 1996.
- CALDER, M. et al. Feature interaction: a critical review and considered forecast. *Computer Networks*, v. 41, p. 115–141, 2003. Disponível em: <http://eprints.gla.ac.uk/2874/1/feature1calder.pdf>.
- CASTILLO, P. A. et al. SOAP vs REST: comparing a master-slave GA implementation. *CoRR*, abs/1105.4978, 2011. Disponível em: <http://arxiv.org/abs/1105.4978>.
- CHANDRAKANTH, S. et al. Internet of things. *International Journal of Innovations and Advancement in Computer Science IJIACS*, v. 3, October 2014.
- DAVIS, J.; GOADRICH, M. The relationship between precision-recall and roc curves. In: *Proceedings of the 23rd International Conference on Machine Learning*. New York, NY, USA: ACM, 2006. (ICML '06), p. 233–240. ISBN 1-59593-383-2. Disponível em: <http://doi.acm.org/10.1145/1143844.1143874>.

- ELIZONDO, D. The linear separability problem: Some testing methods. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, v. 17, n. 2, March 2006. Disponível em: <http://sci2s.ugr.es/keel/pdf/specific/articulo/IEEETNN06.pdf>.
- FIRDAUSI, I. et al. Analysis of machine learning techniques used in behavior-based malware detection. In: *Advances in Computing, Control and Telecommunication Technologies (ACT), 2010 Second International Conference on*. [S.l.: s.n.], 2010. p. 201–203.
- FRANCA, T. et al. Web das coisas: Conectando dispositivos físicos ao mundo digital. In: . Porto Alegre, RS: SBC: [s.n.], 2011. v. 1, p. 103–146.
- GUINARD, D.; TRIFA, V. Towards the web of things: Web mashups for embedded devices. In: *Workshop on Mashups, Enterprise Mashups and Lightweight Composition on the Web, International World Wide Web Conferences*. Madrid, Spain: [s.n.], 2009.
- HALL, M. et al. The weka data mining software: An update. *SIGKDD Explorations*, v. 11, 2009.
- HEFFELFINGER, D. *Java EE 7 with GlassFish 4 Application Server*. Third. [S.l.]: Packt Publishing Ltd, 2014.
- KARTHIKEYAN, T.; THANGARAJU, P. Analysis of classification algorithms applied to hepatitis patients. *International Journal of Computer Applications*, v. 62, n. 15, 2013.
- KERR, A.; HALL, H.; KOZUB, S. *Doing Statistics with SPSS*. [S.l.]: SAGE Publications, 2002. ISBN 0 7619 7384 2.
- KOTSIANTIS, S. Supervised machine learning: A review of classification techniques. *Informatica*, v. 31, p. 249–268, 2007. Disponível em: <http://www.informatica.si/index.php/informatica/article/view/148/140>.
- KRANZ, M.; HOLLEIS, P.; SCHMIDT, A. Embedded interaction interacting with the internet of things. *IEEE INTERNET COMPUTING*, 2010.
- MATERNAGHAN, C.; TURNER, K. J. Policy conflicts in home automation. *Computer Networks*, v. 57, 2013.
- MELVILLE, P.; MOONEY, R. J. Constructing diverse classifier ensembles using artificial training examples. In: *Eighteenth International Joint Conference on Artificial Intelligence*. [S.l.: s.n.], 2003. p. 505–510.
- MELVILLE, P.; MOONEY, R. J. Creating diversity in ensembles using artificial data. *Information Fusion: Special Issue on Diversity in Multiclassifier Systems*, 2004. Submitted.
- METZ, C. Basic principles of roc analysis. *Seminars in Nuclear Medicine*, v. 8, n. 4, p. 283–298, 1978. ISSN 0001-2998. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0001299878800142>.

MICHIE, D.; SPIEGELHALTER, D.; TAYLOR, C. (Ed.). *Machine Learning, Neural and Statistical Classification*. [s.n.], 1994. Disponível em: <http://www1.maths.leeds.ac.uk/~charles/statlog/>.

MINERAUD, J. et al. A gap analysis of internet-of-things platforms. *Computer Communications*, v. 89–90, p. 5 – 16, 2016. ISSN 0140-3664. Internet of Things Research challenges and Solutions. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0140366416300731>.

MULLIGAN, G.; GRACANIN, D. A comparison of soap and rest implementations of a service based interaction independence middleware framework. In: *Proceedings of the 2009 Winter Simulation Conference (WSC)*. [S.l.: s.n.], 2009. p. 1423–1432. ISSN 0891-7736.

NAKAMURA, M. et al. Considering online feature interaction detection and resolution for integrated services in home network system. *Feature Interactions in Software and Communication Systems X*, p. 191–206, 2009.

NAKAMURA, M.; IKEGAMI, K.; MATSUMOTO, S. Considering impacts and requirements for better understanding of environment interactions in home network services. *Computer Networks*, Elsevier, v. 57, p. 2442–2453, 2013.

NHLABATSI, A.; LANEY, R.; NUSEIBEH, B. Feature interaction: the security threat from within software systems. *Progress in Informatics*, n. 5, p. 75–89, 2008. Disponível em: http://www.nii.jp/pi/n5/5_75.pdf.

OLSON, D.; DELEN, D. *Advanced Data Mining Techniques*. [S.l.]: Springer-Verlag Berlin Heidelberg, 2008. ISBN 978-3-540-76916-3.

PAPAZOGLU, M. *Web Services: Principles and Technology*. [S.l.]: Pearson, 2008. ISBN 978-0-321-15555-9.

PAUTASSO, C. Restful web services: Principles, patterns, emerging technologies. In: *Springer Science+Business Media*. New York: [s.n.], 2014. Disponível em: <http://vis.uky.edu/~cheung/courses/ee586/papers/Pautasso2014.pdf>.

PIYARE, R. Internet of things: Ubiquitous home control and monitoring system using android based smart phone. *International Journal of Internet of Things*, Elsevier, p. 5–11, 2013.

QUINLAN, R. *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers, 1993.

ROMAN, R.; ZHOU, J.; LOPEZ, J. On the features and challenges of security and privacy in distributed internet of things. *Computer Networks*, v. 57, n. 10, p. 2266 – 2279, 2013. ISSN 1389-1286. Towards a Science of Cyber Security and Identity Architecture for the Future Internet. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1389128613000054>.

SIEGMUND, N. et al. Predicting performance via automated feature-interaction detection. In: *34th International Conference on Software Engineering (ICSE '12)*. Piscataway, NJ, USA: IEEE Press, 2012. p. 167–177.

SIEGMUND, N. et al. Spl conqueror: Toward optimization of non-functional properties in software product lines. *Software Quality Journal*, v. 20, n. 3, p. 487–517, 2012. ISSN 1573-1367. Disponível em: <http://dx.doi.org/10.1007/s11219-011-9152-9>.

SUNDMAEKER, H. et al. (Ed.). *Vision and Challenges for Realising the Internet of Things*. [S.l.]: European Union, 2010. ISBN 978-92-79-15088-3.

THUM, T. et al. A classification and survey of analysis strategies for software product lines. *ACM Comput. Surv.*, v. 47, May 2014.

WAGH, K.; THOOL, R. A comparative study of soap vs rest web services provisioning techniques for mobile host. *Journal of Information Engineering and Applications*, v. 2, n. 5, 2012.

WANT, R.; SCHILIT, B. N.; JENSON, S. Enabling the internet of things. *IEEE Computer*, Citeseer, v. 48, n. 1, p. 28–35, 2015.

WEBER, R. H.; WEBER, R. *Internet of Things - Legal Perspectives*. [S.l.]: Springer, 2010.

WEISS, M.; ESFANDIARI, B.; LUO, Y. Towards a classification of web service feature interactions. *Computer Networks*, v. 51, p. 359–381, 2007.

WEISS, M.; ORESHKIN, A.; ESFANDIARI, B. Invocation order matters: Functional feature interactions of web services. In: *Proceedings of the First International Workshop on Engineering Service Compositions*. Amsterdam, The Netherlands: [s.n.], 2005. p. 69–76.

WHITMORE, A.; AGARWAL, A.; XU, L. D. The internet of things—a survey of topics and trends. *Information Systems Frontiers*, v. 17, n. 2, p. 261–274, 2015. ISSN 1572-9419. Disponível em: <http://dx.doi.org/10.1007/s10796-014-9489-2>.

WILSON, M.; MAGILL, E.; KOLBERG, M. An online approach for the service interaction problem in home automation. *Consumer Communications and Networking Conference*, IEEE, 2005.

WILSON, M.; MAGILL, E.; KOLBERG, M. Considering side effects in service interactions in home automation - an online approach. In: _____. *Feature Interactions in Software and Communication Systems IX*. [S.l.]: IOS Press, 2008. p. 172–187. ISBN 978-1-58603-845-8.

WITTEN, I.; FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques*. second. [S.l.]: ELSEVIER, 2005. ISBN 0-12-088407-0.

XU, J. et al. Modeling business process of web services with an extended strips operations to detection feature interaction problems runtime. In: IEEE. *Web Services (ICWS), 2011 IEEE International Conference on*. [S.l.], 2011. p. 516–523.

XU, J. et al. Web services feature interaction detection based on situation calculus. In: IEEE. *2010 6th World Congress on Services*. [S.l.], 2010. p. 213–220.

ZASLAVSKY, A. B.; PERERA, C.; GEORGAKOPOULOS, D. Sensing as a service and big data. *CoRR*, abs/1301.0159, 2013. Disponível em: <http://arxiv.org/abs/1301.0159>.

ZHANG, G. Neural networks for classification: A survey. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS*, v. 30, n. 4, November 2000.