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

Apresentação de Monografia

Heron Sanches Gonçalves Pires Ferreira



Universidade Federal da Bahia - Departamento de Ciência da Computação Orientadora: Profa. Dra. Daniela Barreiro Claro Co-Orientador: Roberto Cerqueira Figueiredo Contato: heronsanches@dcc.ufba.br

31 de outubro de 2016

Conteúdo

- Introdução
- 2 Proposta
 - Cenário Levar compras
 - Método de detecção
 - Geração da massa de dados
 - Seleção dos atributos
 - Método ensemble DECORATE
 - Implantação do método proposto
- Validação
 - Método avaliativo stratified-k-fold-cross-validation
 - Experimentos
 - Resultados e discussão
 - Modelos
 - Implantação
- 4 Conclusão e trabalhos futuros

Conteúdo



- Proposta
 - Cenário Levar compras
 - Método de detecção
 - Geração da massa de dados
 - Seleção dos atributos
 - Método ensemble DECORATE
 - Implantação do método proposto
- Validação
 - Método avaliativo stratified-k-fold-cross-validation
 - Experimentos
 - Resultados e discussão
 - Modelos
 - Implantação
- 4 Conclusão e trabalhos futuros



- A Internet está se tornando cada vez mais persistente no cotidiano (Chandrakanth et al., 2014).
- Em 2010 havia aproximadamente 1,5 bilhão de PCs conectados a 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, em 2020 esse número atingirá cerca de 20,8 bilhões.



^{1&}lt;http://www.gartner.com/newsroom/id/3165317>

Introdução

- A Internet está se tornando cada vez mais persistente no cotidiano (Chandrakanth et al., 2014).
- Em 2010 havia aproximadamente 1,5 bilhão de PCs conectados a Internet e mais que 1 bilhão de telefones móveis (Sundmaeker et al., 2010).
- Segundo Gartner¹, 6,4 bilhões de coisas estarão conectadas



Introdução

Introdução

- A Internet está se tornando cada vez mais persistente no cotidiano (Chandrakanth et al., 2014).
- Em 2010 havia aproximadamente 1,5 bilhão de PCs conectados a 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, em 2020 esse número atingirá cerca de 20,8 bilhões.





Coisa: veículo, animal, dispositivo doméstico, pessoa, ...





Imagem: Baseada em (Barrett, 2012)



Internet das Coisas (IoT)

Rede mundial de objetos (coisas) unicamente endereçáveis e interconectados, seguindo os protocolos dos padrões de comunicação (Enterprise et al., 2008).



Imagem: 2





Internet das Coisas (IoT)

Devido ao crescente número de coisas sendo conectadas, surgem diversos desafios, a exemplo de:

- Disponibilidade de uma interface de comunicação (acesso aos serviços e informações dos dispositivos) comum aos objetos.
- Detecção e resolução de efeitos colaterais indesejáveis.



Internet das Coisas (IoT)

Devido ao crescente número de coisas sendo conectadas, surgem diversos desafios, a exemplo de:

- Disponibilidade de uma interface de comunicação (acesso aos serviços e informações dos dispositivos) comum aos objetos.
- Detecção e resolução de efeitos colaterais indesejáveis.



Desafios IoT: interface de comunicação padrão

Disponibilização das coisas como serviços Web RESTFul.

• Mais leve que os serviços Web baseados em SOAP.

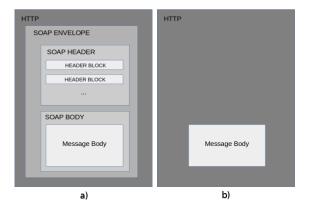


Imagem: Baseada em (Pautasso, 2014)



Desafios IoT: interface de comunicação padrão

Disponibilização das coisas como serviços Web RESTFul.

 Diferentes representações de um mesmo recurso (aumenta a interoperabilidade)

Exemplos

JSON, XML, TEXT, TEXT HTML.

Interface com semântica bem definida.



Desafios IoT: interface de comunicação padrão

Disponibilização das coisas como serviços Web RESTFul.

 Diferentes representações de um mesmo recurso (aumenta a interoperabilidade)

Exemplos

JSON, XML, TEXT, TEXT HTML.

Interface com semântica bem definida.

Exemplos

Utilização dos métodos HTTP (GET, PUT, DELETE, POST, HEAD, OPTIONS, dentre outros).



Desafios IoT: deteccão e resolução de efeitos colaterais indesejáveis

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).



Desafios IoT: detecção e resolução de efeitos colaterais indesejáveis

Quando a composição de *features* leva a algum comportamento não esperado - interação de características, esta pode resultar em efeitos colaterais indesejáveis: um estado inconsistente do sistema, um sistema instável ou dados imprecisos (NHLABATSI et al., 2008).



Introdução Proposta Validação Conclusão e trabalhos futuros

Conteúdo

- Introdução
- 2 Proposta
 - Cenário Levar compras
 - Método de detecção
 - Geração da massa de dados
 - Seleção dos atributos
 - Método ensemble DECORATE
 - Implantação do método proposto
- Validação
 - Método avaliativo stratified-k-fold-cross-validation
 - Experimentos
 - Resultados e discussão
 - Modelos
 - Implantação
- 4 Conclusão e trabalhos futuros



Proposta

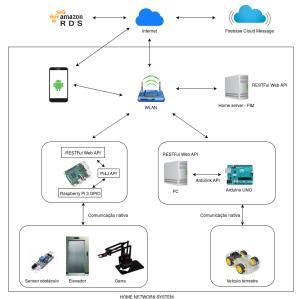
Detectar efeitos colaterais indesejáveis de uma maneira inteligente, utilizando o método *ensemble* DECORATE, no cenário "Levar compras" do *Home Network System* (HNS).



Rede doméstica de coisas (aparelhos domésticos e sensores) com capacidade de conectividade de rede e, interface de controle e monitoramento. Os dispositivos da casa são compostos uns com os outros para prover funcionalidades que atendam aos requisitos dos usuários da casa (Nakamura et al., 2013).



HNS do Cenário Levar compras





Objetos das compras (cestas)

largura	altura	comprimento	massa	diâmetro	nome
θі	0.031	i Θi	0.003	0.009	Refrigerante nº 4 - Soda
0.01	0.014	i 0.03 i	0.001	0	Caixa Pasta de dente - vermelha.
0.004	0.02	0.014	0.001	Ö	Batata
0.004	0.02	0.014	0.001	0	Café
0 i	0.015	i 0 i	0.003	0.008	Café Classic
0.005	0.02	i 0.011 i	0.001	0	Cerveja em lata III
0 j	0.028	i 0i	0.001	0.007	Champanhe - Azul
0.004 j	0.02	j 0.014 j	0.001	0	Farinha de Trigo
0 j	0.03	j Θj	0.001	0.008	Garrafa de Rum
0.012 j	0.016	j 0.004 j	0.004	0	Goiabada
0 j	0.03	j Θj	0.001	0.008	Garrafa de Vodka Special
0.027 j	0.014	j 0.018 j	0.001	0	Pote de sorvete
0 j	0.032	j 0 j	0.002	0.007	Vinho branco
0.012 j	0.03	j 0.006 j	0.002	0	Limpa vidros
0.021	0.015	j 0.007 j	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	Baú 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
Θ	0.03	0	0.003	0.022	Barril em madeira
0.013	0.017	j 0.0063 j	0.002	0	Fermento em pó
0.077	0.0197	j 0.038 j	0.001	Θ	Colchão p/ berço azul
0.01	0.0432	j 0 j	0.001	0	Jogo de esfregões
0.0181	0.0258	j 0.0017 j	0.001	0	Pão de Queijo
0.022	0.0412	0.0135	0.001	0	Rastelo

Imagem: Objetos das compras (cestas). Medidas em metros (m).



Exemplos de Compras (cestas)



Imagem: Exemplos de cestas.



Cenário Levar compras

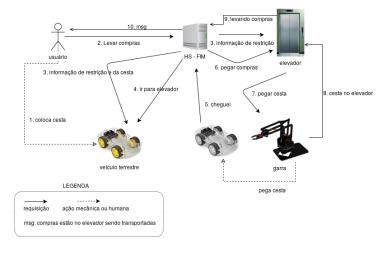


Imagem: Cenário Levar compras.



- Confeccionadas 59 cestas.
- Verificado se as cestas cabiam dentro do elevador.
- Registro no banco de dados relacional na nuvem.
- Cenário Levar compras executado 59 vezes.



- Confeccionadas 59 cestas.
- Verificado se as cestas cabiam dentro do elevador.
- Registro no banco de dados relacional na nuvem
- Cenário Levar compras executado 59 vezes.



- Confeccionadas 59 cestas.
- Verificado se as cestas cabiam dentro do elevador.
- Registro no banco de dados relacional na nuvem.
- Cenário Levar compras executado 59 vezes.



- Confeccionadas 59 cestas.
- Verificado se as cestas cabiam dentro do elevador.
- Registro no banco de dados relacional na nuvem.
- Cenário Levar compras executado 59 vezes.



- Algumas vezes o comportamento da garra fazia com que a cesta n\u00e3o ficasse corretamente no compartimento do elevador, impedindo o elevador continuar com a funcionalidade "Levar compras". Provocando um efeito colateral indesej\u00e1vel.
- Cada cesta foi rotulada como "É efeito colateral" ou "não é éfeito colateral" e atualizado valor no banco.



- Algumas vezes o comportamento da garra fazia com que a cesta não ficasse corretamente no compartimento do elevador, impedindo o elevador continuar com a funcionalidade "Levar compras". Provocando um efeito colateral indesejável.
- Cada cesta foi rotulada como "É efeito colateral" ou "não é éfeito colateral" e atualizado valor no banco.



- Uma cesta com 10 objetos tem $10 \times 5 = 50$ atributos.
- Usado os valores das médias e totais da medidas, assim como a quantidade de objetos. Reduzindo a 11 atributos.



- Uma cesta com 10 objetos tem $10 \times 5 = 50$ atributos.
- Usado os valores das médias e totais da medidas, assim como a quantidade de objetos. Reduzindo a 11 atributos.



- Construído arquivo "arff", padrão WEKA (Hall et al., 2009), com 59 instâncias.
- Cada instância representa uma cesta com 11 parâmetros numéricos mais 1 nominal.
- Gerou-se uma visualização do espaço de atributos em 2D, ou seja, par a par.

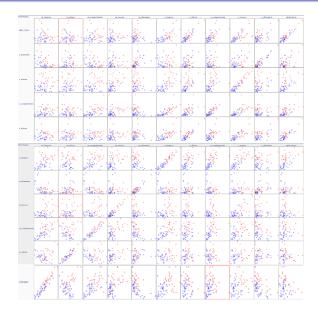


- Construído arquivo "arff", padrão WEKA (Hall et al., 2009), com 59 instâncias.
- Cada instância representa uma cesta com 11 parâmetros numéricos mais 1 nominal.
- Gerou-se uma visualização do espaço de atributos em 2D, ou seja, par a par.



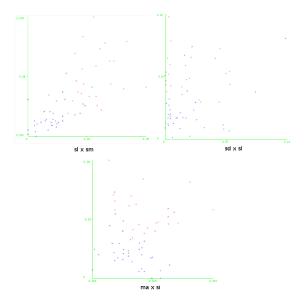
- Construído arquivo "arff", padrão WEKA (Hall et al., 2009), com 59 instâncias.
- Cada instância representa uma cesta com 11 parâmetros numéricos mais 1 nominal.
- Gerou-se uma visualização do espaço de atributos em 2D, ou seja, par a par.







Seleção dos atributos - 3 pares selecionados





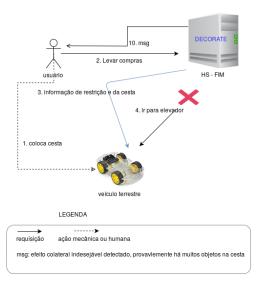
Validação



Parâmetros DECORATE (Melville e Mooney, 2004)



Implantação do método proposto







Introdução Proposta Validação Conclusão e trabalhos futuros

Conteúdo

- - Cenário Levar compras
 - Método de deteccão
 - Geração da massa de dados
 - Seleção dos atributos
 - Método ensemble DECORATE
 - Implantação do método proposto
- Validação
 - Método avaliativo stratified-k-fold-cross-validation
 - Experimentos
 - Resultados e discussão
 - Modelos
 - Implantação



- Verdadeiro positivo (TP): instância que pertence a classe positiva e que foi corretamente classificada como positiva.
- Verdadeiro negativo (TN): instância que pertence a classe
- Falso positivo (FP): instância que pertence a classe negativa e
- Falso negativo (FN): instância que pertence a classe positiva e



- 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
- Falso negativo (FN): instância que pertence a classe positiva e



- 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



- 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 (Metz, 1978)

 Accuracy: taxa correta de classificação em relação a todos os exemplos.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$



Precision, Recall (Davis e Goadrich, 2006)

• Precision: Dentre todos os exemplos classificados como positivos, quais realmente são positivos ?

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

• Recall ou True Positive Rate - TPR: Dentre os exemplos

$$Recall = \frac{TP}{TP + FN} \tag{3}$$



Precision, Recall (Davis e Goadrich, 2006)

 Precision: Dentre todos os exemplos classificados como positivos, quais realmente são positivos ?

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

 Recall ou True Positive Rate - TPR: 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?

$$Recall = \frac{TP}{TP + FN} \tag{3}$$



stratified-k-fold-cross-validation

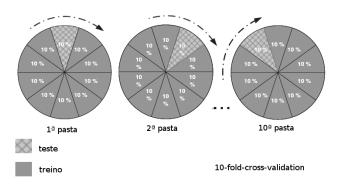


Imagem: k-fold-cross-validation. Adaptado de (Olson e Delen, 2008).



- Para cada par de atributos selecionados utilizou-se stratified-10-fold-cross-validation (repetido dez vezes) juntamente com o ensemble DECORATE a fim de verificar se é possível generalizar pelo menos um modelo.
- Em cada verificação construi-se um modelo com diferentes
- Selecionou-se a curva que mais crescia e se mantia regular e,
- Verificou a partir de que ponto esta curva tinha os melhores
- Realizou um procedimento de geração de modelo final a partir



- Para cada par de atributos selecionados utilizou-se stratified-10-fold-cross-validation (repetido dez vezes) juntamente com o ensemble DECORATE a fim de 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 e, construi-se uma curva de aprendizado.
- Selecionou-se a curva que mais crescia e se mantia regular e, com melhor taxa de recall.
- Verificou a partir de que ponto esta curva tinha os melhores valores de recall, precision e accuracy.
- Realizou um procedimento de geração de modelo final a partir desse melhor ponto.



- Para cada par de atributos selecionados utilizou-se stratified-10-fold-cross-validation (repetido dez vezes) juntamente com o ensemble DECORATE a fim de 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 e, construi-se uma curva de aprendizado.
- Selecionou-se a curva que mais crescia e se mantia regular e, com melhor taxa de recall.
- Verificou a partir de que ponto esta curva tinha os melhores valores de recall, precision e accuracy.
- Realizou um procedimento de geração de modelo final a partir desse melhor ponto.



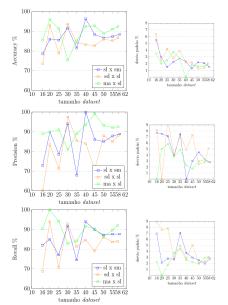
- Para cada par de atributos selecionados utilizou-se stratified-10-fold-cross-validation (repetido dez vezes) juntamente com o ensemble DECORATE a fim de verificar se é possível generalizar pelo menos um modelo.
- Em cada verificação construi-se um modelo com diferentes frações do dataset, variando de 16 instâncias a 59 instâncias e, construi-se uma curva de aprendizado.
- Selecionou-se a curva que mais crescia e se mantia regular e, com melhor taxa de *recall*.
- Verificou a partir de que ponto esta curva tinha os melhores valores de recall, precision e accuracy.
- Realizou um procedimento de geração de modelo final a partir



- Para cada par de atributos selecionados utilizou-se stratified-10-fold-cross-validation (repetido dez vezes) juntamente com o ensemble DECORATE a fim de 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 e, construi-se uma curva de aprendizado.
- Selecionou-se a curva que mais crescia e se mantia regular e, com melhor taxa de recall.
- Verificou a partir de que ponto esta curva tinha os melhores valores de recall, precision e accuracy.
- Realizou um procedimento de geração de modelo final a partir desse melhor ponto.



Resultados: 10×stratified-10-fold-cross-validation





Resultados: curva de aprendizado selecionada

nº i	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
P, DP	93.5, 2.29	99.17, 1.71	93.17, 2.93	92.17, 3.25	92.5, 2
R, DP	91.69, 2.56	90.25, 2.21	86.33, 1.91	89.71, 2.97	92.05, 3

Tabela: Valores das curvas "ma x sl" da Figura 12 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 %.

$$(accuracy \ge (92.75-1.78)) \land (precision \ge (99.17-1.71)) \land (recall \ge 90.25)$$

$$(4)$$



Resultados: geração do modelo final

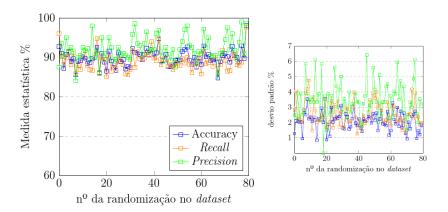


Imagem: stratified-10-fold-cross-validation (repetido dez vezes). 45 instâncias foram escolhidas randomicamente em cima do dataset de 59 instâncias. Procedimento repetido até que satisfizesse a expressão 4



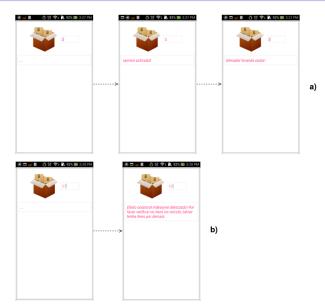
Resultados: geração do modelo final

		Desvio Padrão %
Accuracy %	98.05	1.81
Precision %	98.5	3.2
Recall %	97.67	2

Tabela: Valores da validação do modelo final.



Resultados: Implantação do modelo final





Conteúdo

- Introdução
- 2 Proposta
 - Cenário Levar compras
 - Método de detecção
 - Geração da massa de dados
 - Seleção dos atributos
 - Método ensemble DECORATE
 - Implantação do método proposto
- Validação
 - Método avaliativo stratified-k-fold-cross-validation
 - Experimentos
 - Resultados e discussão
 - Modelos
 - Implantação
- Conclusão e trabalhos futuros



- Foi construído um HNS com o cenário "Levar compras".
- A composição de dispositivos provocou efeitos colaterais
- Obteve-se um modelo baseado no ensemble DECORATE com
- Modelo foi implantando no FIM do HNS, desta forma ficou apto a realizar detecção inteligente de efeitos colaterais



- Foi construído um HNS com o cenário "Levar compras".
- A composição de dispositivos provocou efeitos colaterais indesejáveis no cenário "Levar compras".
- Obteve-se um modelo baseado no ensemble DECORATE com
- Modelo foi implantando no FIM do HNS, desta forma ficou apto a realizar detecção inteligente de efeitos colaterais



- Foi construído um HNS com o cenário "Levar compras".
- A composição de dispositivos provocou efeitos colaterais indesejáveis no cenário "Levar compras".
- Obteve-se um modelo baseado no ensemble DECORATE com alto grau de generalização em um dataset de treino com 45 instâncias com os valores com Accuracy = 98.05%, Precision = 98.5% e Recall = 97.67%.
- Modelo foi implantando no FIM do HNS, desta forma ficou apto a realizar detecção inteligente de efeitos colaterais



- Foi construído um HNS com o cenário "Levar compras".
- A composição de dispositivos provocou efeitos colaterais indesejáveis no cenário "Levar compras".
- Obteve-se um modelo baseado no ensemble DECORATE com. alto grau de generalização em um dataset de treino com 45 instâncias com os valores com Accuracy = 98.05%, Precision = 98.5% e Recall = 97.67%.
- Modelo foi implantando no FIM do HNS, desta forma ficou apto a realizar detecção inteligente de efeitos colaterais indesejáveis.



Trabalhos futuros



Referências I

BARRETT, J. The Internet of Things- where the Web and the physical world will meet. TED x CIT, x =independently organized TED event. 2012. Disponível em: https://www.youtube.com/watch?v=QaTlt1C5R-M. 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>.

CHANDRAKANTH, S. et al. Internet of things. International Journal of Innovations and Advancement in Computer Science IJIACS, v. 3, October 2014.



Referências II

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>.

🗐 ENTERPRISE, I. D. N.; MICRO, R. I. G.; NANOSYSTEMS. Internet of Things in 2020: Roadmap for the Future. 2008.

🗐 HALL, M. et al. The weka data mining software: An update. SIGKDD Explorations, v. 11, 2009.

MELVILLE, P.; MOONEY, R. J. Creating diversity in ensembles using artificial data. Information Fusion: Special Issue on Diversity in Multiclassifier Systems, 2004. Submitted.



Referências III

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.

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.I.]: Springer-Verlag Berlin Heidelberg, 2008. ISBN 978-3-540-76916-3.



Referências IV

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>.

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

