

RANQUEAMENTO EXPLORANDO APRENDIZAGEM DE MÁQUINA NA DESCOBERTA DE RECURSOS NA INTERNET DAS COISAS

FELIPE CAMARGO GRUENDEMANN¹; JUAN BURTET²; ANA MARILZA
PERNAS³, ADENAUER CORRÊA YAMIN⁴

¹Universidade Federal de Pelotas – fcgruendemann@inf.ufpel.edu.br

²Universidade Federal de Pelotas – jburtet@inf.ufpel.edu.br

³Universidade Federal de Pelotas – marilza@inf.ufpel.edu.br

⁴Universidade Federal de Pelotas – adenauer@inf.ufpel.edu.br

1. INTRODUÇÃO

A Internet das Coisas (*IoT – Internet of Things*), é caracterizada por ser um ambiente que apresenta uma grande quantidade de objetos, os quais efetuam conexões transitórias na Internet e resultam em um número elevado de serviços disponibilizados (PERERA, 2017). Este cenário demonstra a relevância do desenvolvimento de estudos e pesquisas para cooperar com os mecanismos de seleção de recursos visando uma alta escalabilidade em ambientes de constituição heterogênea (BASTOS, 2014).

No cenário atual existem mais de seis bilhões de objetos conectados à Internet e constantemente fornecendo serviços aos clientes, havendo a previsão de alcançar vinte bilhões até 2020 (GARTNER, 2015). Na medida em que as “coisas” aprovisionem uma ou mais funcionalidades, a abundância de serviços cresce. Tornando uma tarefa difícil e demorada a seleção adequada de recursos que atendam às demandas exigidas dentre uma grande quantidade de recursos.

Para realizar a seleção do melhor recurso nesta infraestrutura provida pela Internet, o processo de descoberta de recursos precisa analisar tanto os requisitos funcionais quanto os não-funcionais. Assim, após a devida identificação e localização dos recursos, é necessário classificá-los em ordem de Qualidade de Serviço (QoS).

A aprendizagem de máquina tem como objetivo entender a lógica dos programas para aprimorar a performance das máquinas em determinadas tarefas, através de experiência prévia (SUCHITHRA, 2015). Deste modo, os algoritmos de aprendizagem revelam-se muito proveitosos em diversos setores, imediatamente têm sido notáveis principalmente na resolução de problemas de mineração de dados, em que massivos grupos de dados que poderão armazenar, de maneira implícita, informações relevantes e que poderão ser descobertas automaticamente. Da mesma forma, estes algoritmos são indispensáveis em esferas de difícil compreensão, em que os próprios humanos não possuem capacidade para o desenvolvimento de algoritmos eficazes.

Considerando todos esses fatores, tomou-se como objetivo central deste trabalho explorar alternativas para contribuir com a arquitetura de software do mecanismo de descoberta de recursos do middleware EXEHDA, em desenvolvimento no grupo de pesquisa, com foco no aprendizado de máquina.

2. METODOLOGIA

Primeiramente foi realizada uma revisão bibliográfica dos conceitos sobre Internet das Coisas, Descoberta e Ranqueamento de Recursos, Middleware EXEHDA e Algoritmos de Árvore de Decisão para a sistematização do conteúdo.

Na continuidade foram analisados alguns algoritmos de Árvores de Decisão aplicados sobre um *dataset* de recursos usando a ferramenta de *Data Mining*

WEKA (WITTEN et al., 2016). O objetivo da análise foi verificar a acurácia dos algoritmos na classificação de serviço considerando os valores de atributos de QoS, bem como examinar o tempo de construção de modelo de cada algoritmo selecionado. A escolha dos algoritmos de *Machine Learning* foi baseada em DILLI et al. (2017).

3. RANQUEAMENTO DE RECURSOS EXPLORANDO ÁRVORES DE DECISÃO

Existem variados tipos de algoritmos de Aprendizagem de Máquina que podem ser classificados de acordo com sua forma de aprendizagem. Para o contexto da Descoberta e Ranqueamento de Recursos, onde o problema é, baseado em um conjunto de atributos de Qualidade de Serviço (QoS), decidir qual o melhor recurso disponível classificando-os de acordo com suas características, é importante escolher um algoritmo que se adeque à essas necessidades de classificação.

Como um algoritmo de aprendizagem pode ser classificado de acordo com o tipo de aprendizagem que ele realiza, uma árvore de decisão pode se encaixar nos seguintes tipos de Aprendizagem:

- Indutiva: a aprendizagem indutiva é aquela que aprende uma função ou regra geral a partir de determinado par de entrada-saída;
- Supervisionada: na aprendizagem supervisionada o agente aprende uma função que faz o mapeamento da entrada para a saída através da observação dos pares de entrada-saída dados como exemplo.
- Classificação: a aprendizagem de classificação é caracterizada por possuir um conjunto finito de saídas possíveis e essas já são conhecidas. Quando este conjunto possui apenas dois valores ele é chamado de classificação Booleana ou binária.

Identificou-se então que os algoritmos de Árvore de Decisão são uma opção muito conveniente para resolver esse problema de classificação de recursos. Porque estes algoritmos são um tipo de aplicação comum para problemas que envolvem a classificação baseada em decisão e aprendizado adaptativo em um conjunto de treinamento. A árvore de decisão é uma solução bem conhecida para implementar essas táticas. Além disso é uma ferramenta de modelagem de decisão que mostra graficamente o processo de classificação de uma determinada entrada para determinados rótulos de classe de saída (WITTEN, 2011).

3.1 ANÁLISE DOS ALGORITMOS NA FERRAMENTA WEKA

Utilizando a ferramenta WEKA¹ foram analisados alguns dos algoritmos de árvore de decisão que são disponíveis no software: FT, J48, LMT, RandomForest, RandomTree, REPTree, SimpleCart. O objetivo da avaliação foi verificar a acurácia dos algoritmos na classificação de serviços considerando os valores de atributos de QoS.

A classificação dos serviços foi realizada pelos algoritmos para classificar o atributo “*Service Classification*” dentre outros cinco atributos do conjunto de dados analisado. Para a realização dos testes, não foram alteradas as configurações padrão da ferramenta WEKA. Os testes foram executados sobre *datasets* de 800, 900 e 1000 recursos para cada um dos algoritmos mencionados.

¹ <https://www.cs.waikato.ac.nz/ml/weka/>

Os resultados dos testes podem ser vistos na tabela de avaliação dos algoritmos.

Tabela 1 - Avaliação dos Algoritmos

Algoritmo	Nº de Instâncias	Tempo de const. de modelo	Acertos	Erros	Acurácia
LMT	800	2,02s	707	93	88,37%
	900	2,29s	798	102	88,66%
	1000	2,89s	882	118	88,2%
	Acurácia média:				88,41%
Random Forest	800	0,79s	695	105	86,87%
	900	0,48s	789	111	87,66%
	1000	0,49s	896	104	89,6%
	Acurácia média:				88,04%
Random Tree	800	0,1s	634	166	79,25%
	900	0,1s	724	176	80,44%
	1000	0,1s	827	173	82,7%
	Acurácia média:				80,79%
J48	800	0,02s	654	146	81,75%
	900	0,015s	715	185	79,44%
	1000	0,02s	848	152	84,8%
	Acurácia média:				81,99%
SimpleCart	800	0,21s	631	169	78,87%
	900	0,075s	728	172	80,88%
	1000	0,07s	842	158	84,2%
	Acurácia média:				81,31%
REPTree	800	0,02s	614	186	76,75%
	900	0,025s	691	209	76,77%
	1000	0,015s	816	184	81,6%
	Acurácia média:				78,37%
FT	800	0,14s	693	107	86,62%
	900	0,15s	797	103	88,55%
	1000	0,19s	883	117	88,3%
	Acurácia média:				87,82%

A Tabela 1, acima, especifica para cada algoritmo analisado: o número de instâncias presentes no *dataset*, o tempo de construção de modelo, número de acertos, número de erros e a acurácia para cada teste realizado. Apresenta também a média da acurácia que cada algoritmo obteve nos testes.

Nesse sentido pode se observar que o algoritmo com a melhor média de acurácia foi o LMT. Contudo esse foi também o algoritmo com o mais elevado tempo de construção de modelo. Outros algoritmos que obtiveram acurácia elevada foram o Random Forest e o FT, mantendo um tempo de construção de modelo bastante inferior.

4. CONCLUSÕES

A contribuição central desse trabalho foi explorar algoritmos de Árvore de Decisão, tendo como perspectiva seu emprego no Ranqueamento de Recursos na IoT para contribuição na arquitetura de software do mecanismo de descoberta de recursos do middleware EXEHDA.

Foi observado que o conjunto bibliográfico apresentado demonstra que o cenário da Internet das Coisas tende a crescer nos próximos anos elevando a quantidade de recursos e serviços disponíveis. Consequentemente o processo de descoberta e ranqueamento desses recursos enfrenta vários desafios para a classificação dos mesmos. Somando esses fatores ao cenário do middleware EXEHDA foi percebida a importância de buscar métodos eficientes de ranqueamento de recursos para esta ferramenta.

A eficiência dos algoritmos de Árvore de Decisão aplicados a datasets que denotam o cenário da IoT para o Ranqueamento de Recursos foi dada através dos testes feitos. E ficou evidente que os algoritmos LMT e Random Forest possuem uma acurácia bastante satisfatória. Sendo que o LMT teve um maior tempo de execução e acurácia média superior se comparado ao Random Forest.

Dessa maneira é possível afirmar que Algoritmos de Árvores de decisão são uma proposta relevante para incorporação no Ranqueamento de Recursos do middleware EXEHDA.

5. REFERÊNCIAS BIBLIOGRÁFICAS

DILI, R.; FILHO, H.; PERNAS, A.; YAMIN, A. EXEHDA-RR: Machine Learning and MCDA with Semantic Web in IoT Resources Classification. **WebMedia '17**, Gramado, Brasil, 2017.

GARTNER. **Gartner says 6.4 billion connected “things” will be in use in 2016, up 30 percent from 2015**. Stamford, Conn., 10 Novembro, 2015. Acessado em 22 set. 2017. Online. Disponível em: <http://www.gartner.com/newsroom/id/3165317>

JR, N. V. D. A.; BASTOS, D. B. Web of Things: Automatic Publishing and Configuration of Devices. In **Proceedings of the 20th Brazilian Symposium on Multimedia and the Web (WebMedia '14)**, New York, USA, p. 67-74, 2014.

PERERA, C. **Sensing as a Service for Internet of Things: A Roadmap**, Leanpub Publishers, 2017.

SUCHITHRA, M; RAMAKRISHNAN M. A Survey on Different Web Service Discovery Techniques. **Indian Journal of Science and Technology**. v.8, p.1-5, 2015.

WITTEN, I. H.; FRANK E.; HALL, M. A. **The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"**, Morgan Kaufmann, 2016.

WITTEN, I. H.; FRANK, E.; HALL, M. A.; CUNNINGHAM, J.; HUMPHREY, M.; et al. **Data Mining: Practical Machine Learning Tools and Techniques**, Morgan Kaufmann, 2011.