

Avaliação do Emprego de Algoritmos de Árvores de Decisão para a Classificação de Recursos na IoT

Felipe Camargo Gruendemann, Juan Burtet, Renato Dilli,
Ana Marilza Pernas, Adenauer Yamin

¹Laboratory of Ubiquitous and Parallel Systems – UFPEL
Pelotas, RS - Brasil

{fcgruendemann, jburtet, renato.dilli, marilza, adenauer}
@inf.ufpel.edu.br

Resumo. *A IoT é um ambiente de alta escalabilidade que disponibiliza uma quantidade crescente de recursos. Nesse cenário, a tarefa de descoberta e seleção destes recursos é um desafio em aberto. Este artigo apresenta uma alternativa para o processo de classificação dos recursos descobertos utilizando algoritmos de Árvores de Decisão. Os testes realizados com a ferramenta WEKA caracterizaram que estes algoritmos podem atingir bons níveis de acurácia.*

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 [de Andrade Junior et al. 2014].

No cenário atual existem mais de seis bilhões de objetos conectados à Internet e constantemente fornecendo serviços aos clientes, havendo a previsão de ultrapassar 100 bilhões até 2025 [BIS Research 2017]. Na medida em que as “coisas” aprovacionem 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 and Ramakrishnan 2015]. Deste modo, as abordagens de aprendizagem revelam-se oportunas em diversos setores, principalmente na resolução de problemas relacionados a mineração de dados.

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. Classificação 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 Classificação 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 possua a melhor acurácia.

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 conveniente para a classificação de recursos. Esta situação é consequência de tais algoritmos serem adequados 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, e 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 et al. 2016].

3. Avaliações Realizadas

A revisão bibliográfica realizada nesse trabalho apontou 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 classificação desses recursos enfrenta vários desafios. Considerando a perspectiva de emprego do *middleware* EXEHDA na IoT, ficou caracterizada a importância de buscar métodos eficientes para seleção de recursos para o *middleware*.

A metodologia adotada para o trabalho foi a análise de alguns algoritmos de Árvores de decisão. Os algoritmos analisados foram aplicados sobre o *dataset Quality of Web Service* (QWS) [Al-Masri and Mahmoud 2007] usando a ferramenta de *Data Mining* WEKA¹. O objetivo da análise foi verificar a acurácia dos algoritmos na classificação de serviços 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].

Utilizando a ferramenta WEKA foram analisados alguns dos algoritmos de árvore de decisão que o software dispõe: FT, J48, LMT, RandomForest, RandomTree, SimpleCart. A classificação dos serviços foi realizada pelos algoritmos a fim de classificar o

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

Tabela 1. Avaliação dos Algoritmos

Algoritmo	Instâncias	Tempo	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%
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%

atributo “Service Classification” dentre outros cinco atributos do conjunto de dados analisado. Para execução dos testes não foram alteradas as configurações padrão da ferramenta WEKA. E estes foram realizados sobre *datasets* de 800, 900 e 1000 recursos para cada um dos algoritmos mencionados. Os resultados dos testes podem ser vistos na tabela Avaliação dos Algoritmos.

A Tabela 1 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. A Tabela 1 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 comparado ao LMT.

Para a realização de todos os testes foi utilizada uma máquina com processador Intel Core i7-4510U CPU @ 2,00GHz e memória RAM DDR3 de 8GB.

4. Considerações Finais

A contribuição central desse trabalho foi explorar algoritmos de Árvore de Decisão, tendo como perspectiva seu emprego na Classificação de Recursos na IoT para contribuição na arquitetura de software do mecanismo de descoberta de recursos do *middleware* EXEHDA.

Os testes com os algoritmos de Árvore de Decisão aplicados à classificação de recursos apontaram uma acurácia média acima de 80%, o que por se tratar de uma pré-classificação atende as demandas do grupo referentes a otimização de desempenho do *middleware* quanto ao ranqueamento de recursos [Dilli et al. 2017]. Também ficou caracterizado que os algoritmos LMT e Random Forest foram os que apresentaram melhores resultados. 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 constituem uma alternativa a ser considerada para emprego no serviço de Classificação de Recursos do *middleware* EXEHDA.

A proposta para trabalhos futuros é realizar outros testes envolvendo conjuntos de dados maiores explorando para isto um gerador de *datasets*, e também avaliar outras alternativas de Aprendizagem de Máquina, como os *Support Vector Machines* (SVMs).

Referências

- Al-Masri, E. and Mahmoud, Q. H. (2007). QoS-based discovery and ranking of Web services. In *Proceedings - International Conference on Computer Communications and Networks, ICCCN*, pages 529–534.
- BIS Research (2017). Global Sensors in Internet of Things (IoT) Devices Market, Analysis & Forecast: 2016 to 2022. Technical report.
- de Andrade Junior, N. V., Bastos, D. B., and Prazeres, C. V. S. (2014). Web of Things: Automatic Publishing and Configuration of Devices. *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web*, pages 67–74.
- Dilli, R., Filho, H. K., Pernas, A. M., and Yamin, A. (2017). EXEHDA-RR: Machine Learning and MCDA with Semantic Web in IoT Resources Classification. *WebMedia '17, October 17–20, 2017, Gramado, Brazil*.
- Perera, C. (2017). Sensing as a Service for Internet of Things: A Roadmap.
- Suchithra, M. and Ramakrishnan, M. (2015). A survey on different web service discovery techniques. *Indian Journal of Science and Technology*, 8(15):1–5.
- Witten, I. H., Frank, E., and Hall, M. A. (2016). The WEKA Workbench. *Morgan Kaufmann, Fourth Edition*, pages 553–571.