
LEGAL-Tree: Um algoritmo genético
multi-objetivo lexicográfico para indução de
árvores de decisão

Márcio Porto Basgalupp

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

LEGAL-Tree: Um algoritmo genético multi-objetivo lexicográfico para indução de árvores de decisão

Márcio Porto Basgalupp

Orientador: *Prof. Dr. André Carlos Ponce de Leon Ferreira de Carvalho*

Co-orientador: *Prof. Dr. Alex Alves Freitas*

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP como parte dos requisitos necessários à obtenção para do título em Doutor em Ciências de Computação e Matemática Computacional.

USP - São Carlos
fevereiro/2010

Agradecimentos

À minha mãe Neida Mariza Veiga Porto e ao meu pai Jorge Luiz Basgalupp (*in memoriam*), responsáveis diretos por cada uma das minhas realizações.

Às minhas irmãs Ana Luiza, Suélen e Taiane, e à minha vó Eulina.

Ao meu afilhado Diogo, que é o gremista mais lindo que eu já vi.

Aos meus demais familiares e amigos, simplesmente por existirem e fazerem parte da minha vida.

Aos professores João Baptista da Silva, João Artur de Souza, Gertrudes Dandolini, Gil Carlos Medeiros, Karin Becker e Duncan Ruiz, os quais levarei sempre comigo por participarem efetivamente da minha vida acadêmica e pessoal.

A todos os colegas do LABIC e BIOCOMP, do ICMC-USP, e do GPIN, da PUCRS, com os quais compartilhei ótimos momentos e que, direta ou indiretamente, influenciaram na realização deste trabalho.

A todos os funcionários e professores do ICMC-USP por me proporcionarem o excelente ambiente de trabalho que tive durante todo o tempo do doutorado.

Ao colega e amigo Rodrigo Barros, que também participou efetivamente do desenvolvimento desta pesquisa.

Ao meu co-orientador, Prof. Dr. Alex Alves Freitas, não apenas pela dedicação na realização do meu trabalho, mas também por ser extremamente atencioso e prestativo durante o tempo em que estive na Inglaterra.

Ao meu orientador, Prof. Dr. André Carlos Ponce de Leon Ferreira de Carvalho, o qual é uma das pessoas mais gentis que já conheci na minha vida. Se eu terminar a minha carreira com a metade de seu currículo, serei realizado profissionalmente. Mas, se eu terminar a vida com a metade de sua humildade, serei uma pessoa realizada.

Aos autores do algoritmo GALE, Xavier Llorca and Josep Garrell, por gentilmente me fornecerem esse algoritmo.

À FAPESP, pelo financiamento deste trabalho.

Resumo

Dentre as diversas tarefas em que os algoritmos evolutivos têm sido empregados, a indução de regras e de árvores de decisão tem se mostrado uma abordagem bastante atrativa em diversos domínios de aplicação. Algoritmos de indução de árvores de decisão representam uma das técnicas mais populares em problemas de classificação. Entretanto, os algoritmos tradicionais de indução apresentam algumas limitações, pois, geralmente, usam uma estratégia gulosa, *top down* e com particionamento recursivo para a construção das árvores. Esses fatores degradam a qualidade dos dados, os quais podem gerar regras estatisticamente não significativas. Este trabalho propõe o algoritmo LEGAL-Tree, uma nova abordagem baseada em algoritmos genéticos para indução de árvores de decisão. O algoritmo proposto visa evitar a estratégia gulosa e a convergência para ótimos locais. Para isso, esse algoritmo adota uma abordagem multi-objetiva lexicográfica. Nos experimentos realizados sobre bases de dados de diversos problemas de classificação, a função de *fitness* de LEGAL-Tree considera as duas medidas mais comuns para avaliação das árvores de decisão: acurácia e tamanho da árvore. Os resultados obtidos mostraram que LEGAL-Tree teve um desempenho equivalente ao algoritmo SimpleCart (implementação em Java do algoritmo CART) e superou o tradicional algoritmo J48 (implementação em Java do algoritmo C4.5), além de ter superado também o algoritmo evolutivo GALE. A principal contribuição de LEGAL-Tree não foi gerar árvores com maior acurácia preditiva, mas sim gerar árvores menores e, portanto, mais compreensíveis ao usuário do que as outras abordagens, mantendo a acurácia preditiva equivalente. Isso mostra que LEGAL-Tree obteve sucesso na otimização lexicográfica de seus objetivos, uma vez que a idéia era justamente dar preferência às árvores menores (em termos de número de nodos) quando houvesse equivalência de acurácia.

Abstract

Among the several tasks evolutionary algorithms have been successfully employed, the induction of classification rules and decision trees has been shown to be a relevant approach for several application domains. Decision tree induction algorithms represent one of the most popular techniques for dealing with classification problems. However, conventionally used decision trees induction algorithms present limitations due to the strategy they usually implement: recursive top-down data partitioning through a greedy split evaluation. The main problem with this strategy is quality loss during the partitioning process, which can lead to statistically insignificant rules. In this thesis we propose the LEGAL-Tree algorithm, a new GA-based algorithm for decision tree induction. The proposed algorithm aims to prevent the greedy strategy and to avoid converging to local optima. For such, it is based on a lexicographic multi-objective approach. In the experiments which were run in several classification problems, LEGAL-Tree's fitness function considers two of the most common measures to evaluate decision trees: accuracy and tree size. Results show that LEGAL-Tree performs similarly to SimpleCart (CART Java implementation) and outperforms J48 (C4.5 Java implementation) and the evolutionary algorithm GALE. LEGAL-Tree's main contribution is not to generate trees with the highest predictive accuracy possible, but to provide smaller (and thus more comprehensible) trees, keeping a competitive accuracy rate. LEGAL-Tree is able to provide both comprehensible and accurate trees, which shows that the lexicographic fitness evaluation is successful since its goal is to prioritize smaller trees (fewer number of nodes) when there is equivalency in terms of accuracy.

Gracias por visitar este Libro Electrónico

Puedes leer la versión completa de este libro electrónico en diferentes formatos:

- HTML(Gratis / Disponible a todos los usuarios)
- PDF / TXT(Disponible a miembros V.I.P. Los miembros con una membresía básica pueden acceder hasta 5 libros electrónicos en formato PDF/TXT durante el mes.)
- Epub y Mobipocket (Exclusivos para miembros V.I.P.)

Para descargar este libro completo, tan solo seleccione el formato deseado, abajo:

