



DEFESA DE TESE DE DOUTORADO

Programa de Pós-Graduação em Ciência da Computação

UBERBAND: META-APRENDIZADO E OTIMIZAÇÃO BASEADA EM BANDIDOS MULTI-ARMADOS PARA SELEÇÃO EFICIENTE E EFETIVA DE PROCESSOS COMPLETOS

ALUNO: Silvia Cristina Nunes das Dores

ORIENTADOR: Dr. Duncan Dubugras Alcoba Ruiz

CO-ORIENTADOR: Dr. Carlos Manuel Milheiro de Oliveira Pinto Soares

BANCA EXAMINADORA: Dr. André Carlos Ponce de Leon Ferreira de Carvalho (ICMC/USP), Dr. Ricardo Bastos Cavalcante Prudencio (CIn/UFPE), Dr. Felipe Rech Meneguzzi (PPGCC/PUCRS)

DATA: 29 de abril de 2019

LOCAL: Prédio 32, sala 506.2

HORÁRIO: 10:00

RESUMO:

Na medida em que tecnologias para gerenciamento e armazenamento de dados se tornam amplamente disponíveis, torna-se um desafio fornecer aos usuários sistemas eficazes de análise e compreensão desses dados. Descoberta de Conhecimento em Bases de Dados (DCBD) é o processo não-trivial de extração de padrões interessantes, válidos e úteis a partir de dados. Este processo inclui desde a seleção de dados, até a interpretação dos padrões identificados. Especialmente para usuários não especialistas, a definição e gestão de um processo de DCBD são atividades complexas, pois é exigido o conhecimento sobre como escolher as operadores adequados dentre a gama disponível para cada etapa, como configurá-los e como interpretar sua saída. Seleção Automática de Processos Completos (SPC) objetiva auxiliar usuários de DCBD na tarefa onerosa de escolher o processo completo de operadores, que inclui métodos de pré-processamento, algoritmos de aprendizado de máquina e suas configurações de hiper-parâmetros mais adequados a um determinado problema. Embora diversas soluções já existam para esta tarefa, tais soluções são limitadas do ponto de vista da avaliação do processo completo i) algumas soluções não realizam experimentação do processo e se baseiam apenas em estimativas de desempenho de problemas similares, o que pode levar a recomendações não-precisas e ii) as demais soluções avaliam os processos completos repetidas vezes sobre o conjunto de treinamento inteiro até encontrar a melhor opção. Estas últimas soluções geralmente obtêm resultados mais precisos, porém, se tornam computacionalmente custosas em termo de tempo, à medida em que os conjuntos de dados aumentam e novos algoritmos são desenvolvidos. Neste sentido, esta pesquisa propõe e analisa um novo algoritmo para SPC, denominado Uberband, que combina Meta-Aprendizado, para a estimar a probabilidade de amostragem de operadores, e otimização baseada em bandidos multi-armados, para realização de alocação adaptativa de instâncias do conjunto de treinamento durante o processo de otimização. Resultados da análise experimental comparativa com soluções estado-da-arte em SPC, indicaram que Uberband proporciona uma



SPC com desempenho similar e em um tempo expressivamente menor do que as soluções atuais.

Palavras-Chave: aprendizado de máquina automático, meta-aprendizado, otimização baseada em bandidos, seleção de processo completo.