

### III Workshop de Dissertações de Mestrado do PPgSI (2016)

#### FICHA DA PESQUISA<sup>i</sup>

DADOS GERAIS					
<b>Título do projeto de pesquisa</b>	Técnicas de transferência de aprendizagem aplicadas a modelos QSAR para regressão.				
<b>Orientando</b>	Rodolfo da Silva Simões.				
<b>Orientador(es)</b>	Patrícia Rufino Oliveira				
<b>Momento atual</b>	<input type="checkbox"/> 2º semestre	<input type="checkbox"/> 3º semestre	<input checked="" type="checkbox"/> 4º semestre	<input type="checkbox"/> 5º semestre	<input type="checkbox"/> 6º semestre
<b>Qualificação</b>	<input type="checkbox"/> Qualificação já realizada em: ____/____/____		<input checked="" type="checkbox"/> Qualificação planejada para: 24/08/2016		
<b>Defesa</b>	Prazo máximo para depósito: 26/02/2018		Depósito planejado para: 25/02/2017		
<b>Linha e Área de pesquisa</b>	<input type="checkbox"/> Gestão e desenvolvimento de Sistemas: <input type="checkbox"/> BD <input type="checkbox"/> Gestão de SI <input type="checkbox"/> Eng. de Software <input type="checkbox"/> IHC		<input checked="" type="checkbox"/> Inteligência de Sistemas: <input type="checkbox"/> IA <input checked="" type="checkbox"/> Rec. de Padrões <input type="checkbox"/> Proc. Gráfico		
<b>Área de aplicação</b>	<input type="checkbox"/> Ambientes Corporativos <input type="checkbox"/> Bioinformática <input type="checkbox"/> Biometria <input type="checkbox"/> Economia	<input type="checkbox"/> Educação <input type="checkbox"/> Educação a Distância <input type="checkbox"/> Internet <input type="checkbox"/> Jogos	<input type="checkbox"/> Linguagem Natural <input type="checkbox"/> Linguística <input type="checkbox"/> Processos de Negócio <input checked="" type="checkbox"/> Química	<input type="checkbox"/> Redes Sociais <input type="checkbox"/> Robótica <input type="checkbox"/> Saúde	
<b>Publicações</b>	<ul style="list-style-type: none"> <li>• Aceita: Título do artigo: Recent Advances for Handling Imbalance and Uncertainty in Labelling in Medicinal Chemistry Data Analysis. Título da conferência: SAI Computing Conference 2016</li> <li>• Elaboração: Título do artigo: Uma revisão sistemática sobre métodos de classificação supervisionados em conjunto com a técnica de transferência de aprendizagem.</li> </ul>				

DESCRIÇÃO DO PROJETO DE PESQUISA	
<b>Contextualização / motivação</b>	A transferência de aprendizagem utilizando informações de outros modelos QSAR's com mais informações disponíveis para o mesmo alvo biológico, diminui o esforço e o custo do processo para gerar novos modelos.
<b>Problema de pesquisa</b>	Estudos QSAR são muito custosos tendo em vista que a transformação dos descritores químicos em conjuntos de dados numéricos não é uma tarefa trivial. Além disso, grande parte dos compostos químicos não possui rótulos pré-definidos e as amostras que possuem compostos rotulados são muito pequenas.
<b>Objetivo geral da pesquisa</b>	Este projeto de mestrado tem por objetivo geral avaliar a aplicabilidade de duas técnicas de transferência de aprendizagem para construção de modelos QSAR para regressão.
<b>Trabalhos relacionados</b>	<ul style="list-style-type: none"> <li>• ROSEMBAUM, L. et al. <i>Infering multi-target qsar models with taxonomy-based multi-task learning</i>. Os autores propuseram dois diferentes algoritmos multitarefa que podem explorar a semelhança entre vários alvos biológicos para transferir o conhecimento entre conjuntos de dados de Química Medicinal.</li> <li>• GIRSCHICK, T.; RUCKERT, U.; KRAMER, S. <i>Adapted transfer of distance measures for quantitative structure-activity relationships and data-driven selection of source datasets</i> Os autores propuseram um método de seleção de conjuntos de dados de origem de forma dirigida para a aplicação de técnicas de transferência de aprendizagem em conjuntos de dados de Química Medicinal.</li> <li>• PARDOE, D.; STONE, P. <i>Boosting for regression transfer</i>. Os autores propuseram um método para a técnica de transferência de aprendizagem por instância, o objetivo do algoritmo é filtrar e utilizar apenas alguns dados do domínio de origem junto ao conjunto de dados de destino.</li> </ul>
<b>Justificativa e relevância</b>	De maneira geral, abordagens de transferência de aprendizagem têm se mostrado promissoras para combinar conhecimento previamente obtido em tarefas relacionadas em um único modelo preditivo, seja este para classificação, regressão ou agrupamento. Além disso, estudos QSAR são muito custosos, o que torna desejável a aplicação da técnica de transferência de aprendizagem.
<b>Proposta para Solução</b>	(i) Será proposto o método Regressão por Vetores Suporte (SVR) adaptado para a técnica de transferência de aprendizagem por parâmetros; (ii) Será utilizado o método TrAdaBoost; (iii) Será aplicado um algoritmo de seleção de conjuntos de dados de origem; (iv) Os métodos considerados serão comparados com o método SVR
<b>Dados</b>	Será utilizada a base de dados <i>PubChem BioAssay</i> . Os conjuntos de dados que serão considerados no projeto de

	pesquisa são modelos QSAR para regressão. O problema consiste em prever o valor real do nível de atividade biológica de um composto químico em relação a um alvo biológico.
<b>Validação</b>	Será utilizada a técnica de validação cruzada para avaliar o desempenho dos métodos de transferência de aprendizagem, considerando, ainda, conjuntos de dados de origem e destino com diferentes tamanhos. Como métrica de avaliação, será considerada a medida de Erro Quadrático Médio (MSE).
<b>Limitações</b>	Os conjuntos de dados de Química Medicinal apresentam um número pequeno de moléculas (instâncias), o que geralmente diminui a eficácia dos modelos preditivos construídos.
<b>Resultados esperados</b>	Contribuições científicas: Espera-se que os métodos de transferência de aprendizagem melhorem modelos QSAR gerados para regressão por meio de vetores suporte. Contribuições tecnológicas: Será proposto um novo método de Regressão por Vetores Suporte adaptado para a utilização da técnica de transferência de aprendizagem.

<b>MÉTODO DE PESQUISA</b> [para cada item, marque todas as opções adequadas para seu projeto]			
<b>Gênero</b>	<input type="checkbox"/> Pesquisa teórica	<input checked="" type="checkbox"/> Pesquisa prática	<input checked="" type="checkbox"/> Pesquisa empírica <input type="checkbox"/> Pesquisa metodológica
<b>Natureza</b>	<input type="checkbox"/> Pesquisa básica/pura	<input checked="" type="checkbox"/> Pesquisa aplicada	
<b>Objetivo</b>	<input type="checkbox"/> Pesquisa descritiva	<input checked="" type="checkbox"/> Pesquisa exploratória	<input type="checkbox"/> Pesquisa explicativa <input type="checkbox"/> [outro – escrever]
<b>Abordagem</b>	<input checked="" type="checkbox"/> Pesquisa quantitativa	<input type="checkbox"/> Pesquisa qualitativa	<input type="checkbox"/> Pesquisa mista (quali-quant)
<b>Procedimento(s) técnico(s)</b>	<input checked="" type="checkbox"/> Pesquisa experimental <input checked="" type="checkbox"/> Pesquisa bibliográfica <input type="checkbox"/> Pesquisa documental <input type="checkbox"/> Pesquisa <i>ex-post-facto</i> <input type="checkbox"/> Pesquisa de levantamento	<input checked="" type="checkbox"/> Pesquisa com <i>survey</i> <input type="checkbox"/> Estudo de caso <input type="checkbox"/> Pesquisa participante <input type="checkbox"/> Pesquisa-ação <input type="checkbox"/> Pesquisa etnográfica	<input type="checkbox"/> Pesquisa netnográfica <input type="checkbox"/> Teoria fundamentada em dados ( <i>grounded theory</i> ) <input type="checkbox"/> Ciência do projeto ( <i>Design science research</i> ) <input type="checkbox"/> [outro – escrever]
<b>Fonte(s) de dados</b>	<input checked="" type="checkbox"/> pesquisa de laboratório	<input type="checkbox"/> pesquisa de campo	<input type="checkbox"/> pesquisa bibliográfica <input type="checkbox"/> [outro – escrever]
<b>Técnica(s) / Instrumento(s) de coleta de dados</b>	<input checked="" type="checkbox"/> medição <input type="checkbox"/> questionário <input type="checkbox"/> entrevista <input type="checkbox"/> grupos focais	<input type="checkbox"/> formulário <input checked="" type="checkbox"/> <i>benchmark</i>	<input type="checkbox"/> observação (direta / participante) <input type="checkbox"/> diário de campo / notas de campo <input type="checkbox"/> análise documental (ou de artefatos) <input type="checkbox"/> [outro – escrever]
<b>Técnica(s) de análise de dados</b>	Análise quantitativa: <input checked="" type="checkbox"/> Estatística descritiva <input type="checkbox"/> Estatística inferencial <input type="checkbox"/> [outro – escrever]	Análise qualitativa: <input type="checkbox"/> Análise de conteúdo <input type="checkbox"/> Análise do discurso <input type="checkbox"/> [outro – escrever]	

<b>CRONOGRAMA</b> [altere conforme necessário]																																															
	2015												2016												2017																						
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12											
Estudo da literatura		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X																								
Revisão sistemática													X	X	X	X																															
Definição da proposta															X	X	X																														
Execução da proposta																			X	X	X	X	X																								
Coleta/obtenção de dados																			X																												
Análise dos dados																			X	X																											
Validação da proposta																				X	X	X	X																								
Preparo da qualificação													X	X	X	X	X	X	X																												
Exame de qualificação																					X																										
Escrita da dissertação																					X	X	X																								
Escrita de artigo													X	X	X	X	X	X	X	X	X	X	X																								
Depósito da dissertação																																															
Defesa da dissertação																																															

<sup>i</sup> Esta ficha é uma adaptação da usada no “VIII Workshop de Teses e Dissertações em Sistemas de Informação (WTDSI 2015)” realizado como parte do “XI Simpósio Brasileiro de Sistemas de Informação (SBSI 2015)”