

## V Workshop de Dissertações de Mestrado do PPgSI (2018)

### FICHA DE PESQUISA

DADOS GERAIS				
Título do projeto de pesquisa	Limites de surpresa em Sistemas de Recomendação.			
Orientando	Andre Paulino de Lima			
Orientador(es)	Sarajane Marques Peres			
Semestre no curso, na data do workshop	<input type="checkbox"/> 2º semestre	<input type="checkbox"/> 3º semestre	<input type="checkbox"/> 4º semestre	<input checked="" type="checkbox"/> 5º semestre
Qualificação	[x] Qualificação já realizada em: 21/11/2017 <input type="checkbox"/> Realização da qualificação planejada para: não se aplica			
Defesa	Prazo máximo para depósito: 29/01/2019    Realização da defesa planejada para: 10/10/2018			
Linha e Área de pesquisa	Gestão e Desenvolvimento de Sistemas: <input type="checkbox"/> BD <input type="checkbox"/> Engenharia de Software <input type="checkbox"/> Gestão de TI <input type="checkbox"/> IHC		Inteligência de Sistemas: <input checked="" type="checkbox"/> IA <input type="checkbox"/> Processamento Gráfico <input type="checkbox"/> Reconhecimento de Padrões	
Área de aplicação	<input type="checkbox"/> Ambientes corporativos	<input type="checkbox"/> Educação a distância	<input type="checkbox"/> Língua Natural	<input type="checkbox"/> Redes Sociais
	<input type="checkbox"/> Bioinformática	<input type="checkbox"/> Governo eletrônico	<input type="checkbox"/> Linguística	<input type="checkbox"/> Robótica
	<input type="checkbox"/> Biometria	<input type="checkbox"/> Internet	<input type="checkbox"/> Processos de Negócio	<input type="checkbox"/> Saúde
	<input type="checkbox"/> Dispositivos móveis	<input type="checkbox"/> Jogos	<input type="checkbox"/> Quimioinformática	<input checked="" type="checkbox"/> Sistemas de Recomendação
	<input type="checkbox"/> Educação	<input type="checkbox"/> Jogos sérios		
Publicações associadas ao projeto de mestrado	"Limits to Surprise in Recommender Systems", submetido para RecSys'18, mas rejeitado. Preprint disponível no arxiv ( <a href="http://arxiv.org/abs/1807.03905">http://arxiv.org/abs/1807.03905</a> ). Comentários de revisão sendo endereçados para elaboração de novo artigo, a ser submetido para a conferência ECIR'19.			

DESCRIÇÃO DO PROJETO DE PESQUISA	
Contextualização / motivação	Sistemas de Recomendação (SR), em especial àqueles que adotam estratégia de personalização de recomendação baseada em conteúdo, estão sujeitos ao problema da superespecialização (overspecialisation, filter bubble), no qual o usuário passa a receber recomendações de itens que são muito similares aos itens que receberam avaliação positiva do usuário no passado. Como efeito, o usuário é privado de receber recomendações de novos itens que não sejam similares aos já avaliados, mesmo que estes novos itens sejam relevantes.
Problema de pesquisa	Apesar da superespecialização ser um problema conhecido, não há na literatura métodos ou métricas que possam ser empregados para estimar o grau de superespecialização de um SR.
Objetivo geral da pesquisa	Elaborar e validar uma métrica para avaliar o grau de superespecialização de um SR.
Trabalhos relacionados	<p><b>Sobre a relação entre satisfação do usuário e o balanceamento adequado entre objetivos do sistema</b> (acurácia e não-acurácia): Being accurate is not enough: how accuracy metrics have hurt recommender systems (McNee, Riedl, Konstan; 2006); An empirical analysis of design choices in neighbourhood-based collaborative filtering algorithms (Herlocker, Konstan, Riedel; 2002).</p> <p><b>Sobre os modelos matemáticos que descrevem o conceito de surpresa:</b> Evaluating surprise in recommender systems (Kaminskas, Bridge; 2014); Novelty or surprise? (Barto, Mirolli, Baldassarre; 2013).</p> <p><b>Sobre o impacto da escolha do modelo de representação de dados sobre a avaliação de similaridade entre itens (essencial para surpresa):</b> From frequency to meaning: vector space models of semantics (Turney, Pantel; 2010); Features of similarity (Tversky, 1977); Similarity as Inference (Gershman, 2017 – submitted).</p>
Proposta para solução	Assumindo que a superespecialização é expressa por uma redução sistemática da surpresa embutida nas recomendações produzidas por um SR, este projeto adota uma perspectiva na qual a) a surpresa é um recurso finito em um SR, e b) existem limites mínimo e máximo para surpresa embutível em uma recomendação. Esses limites determinam escala (surpresa potencial normalizada) que pode aferir o nível de superespecialização em um SR.
Dados	Este projeto emprega o conjunto de dados Movielens-1M, amplamente citado na literatura de SR, e um conjunto de dados sintético.
Validação	A estratégia de validação da pesquisa é composta de duas etapas. Na primeira etapa, um conjunto de premissas sobre a natureza da interação entre usuário e SR é assumido e, por método dedutivo, o modelo teórico de surpresa potencial é construído. Com base nesse modelo, a métrica/escala de surpresa potencial normalizada é

