

Emiliano Gonçalves de Castro

Proposta para previsão de evasão baseada  
em padrões de acesso de usuários em jogos  
*online*

Tese apresentada à Escola Politécnica  
da Universidade de São Paulo como  
requisito para obtenção do Título de  
Doutor em Engenharia.

Emiliano Gonçalves de Castro

Proposta para previsão de evasão baseada  
em padrões de acesso de usuários em jogos  
*online*

Tese apresentada à Escola Politécnica  
da Universidade de São Paulo como  
requisito para obtenção do Título de  
Doutor em Engenharia.

Área de concentração:  
Engenharia de Controle e Automação  
Mecânica

Orientador:  
Prof. Dr. Marcos de Sales Guerra  
Tsuzuki

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 11 de julho de 2011.

Emiliano Gonçalves de Castro

Prof. Dr. Marcos de Sales Guerra Tsuzuki

### **Ficha Catalográfica**

Castro, Emiliano Gonçalves de

Proposta para previsão de evasão baseada em padrões de acesso de usuários em jogos *online*. São Paulo, 2011. 76 p.

Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos.

1. Previsão de Evasão. 2. Jogos Eletrônicos. 3. Wavelets. I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos. II. Título.

Aos meus pais.

# Agradecimentos

O impulso responsável pela realização deste trabalho veio do ambiente altamente estimulante e desafiador que sempre encontrei junto aos meus colegas da indústria de jogos eletrônicos. Agradeço a todos aqueles que contribuíram, de uma forma ou de outra, para aguçar a minha curiosidade sobre o tema desta pesquisa.

Agradeço especialmente ao Júlio Vieitez, da Level Up! Games, pelas conversas a respeito deste trabalho e pela gentileza em compartilhar os dados brutos utilizados nesta pesquisa. Sem a sua contribuição, nenhuma análise teria sido possível.

Sou grato também aos meus colegas da Escola Politécnica da USP, particularmente ao Prof. Thiago Martins, por todo o apoio nas discussões sobre a metodologia empregada e sobre a pesquisa de uma forma geral; ao Prof. Romero Tori, pelas valiosas sugestões e críticas na fase de qualificação deste trabalho; ao André Sato Kubagawa e à Christiane, pela atenciosa ajuda na revisão do texto.

Finalmente, crédito ao Prof. Marcos Tsuzuki a motivação para todo o meu trabalho como pesquisador até aqui. Sua orientação segura e paciente foi fundamental do primeiro ao último dia.

# Resumo

O mercado de jogos eletrônicos *online* tem crescido em ritmo acelerado nos últimos anos, particularmente a partir do surgimento do modelo de negócio baseado em serviços. Como consequência, as publicadoras destes jogos passaram a compartilhar problemas comuns na área de serviços, como a erosão do lucro causada pela evasão de usuários. Modelos preditivos têm sido utilizados no combate à evasão em mercados como os de telefonia móvel e de cartões de crédito, setores que detêm um grande volume de informações demográficas e econômicas a respeito dos seus consumidores. Já os publicadores de jogos muitas vezes só possuem o endereço eletrônico dos jogadores.

O objetivo deste trabalho é propor um modelo de previsão de evasão com base exclusivamente nos padrões de acesso de usuários em jogos *online*, onde estes registros temporais são submetidos a um conjunto de operadores que analisam os dados no domínio do plano tempo-frequência, utilizando a Transformada Discreta de Wavelet. Sua principal contribuição está na proposta de parametrização dos dados de entrada para classificadores probabilísticos baseados no algoritmo *k-Nearest Neighbors*.

Testados com dados reais de acessos de usuários ao longo de alguns meses em um jogo *online*, os classificadores foram avaliados com o uso de curvas ROC (*Receiver Operating Characteristic*) e de elevação. A abordagem proposta nesta tese, baseada na análise no domínio do plano tempo-frequência, apresentou resultados satisfatórios. Não apenas superiores se comparados com as abordagens no domínio do tempo ou da frequência, mas também comparáveis aos desempenhos encontrados por modelos com centenas de variáveis preditivas utilizados em outros mercados.

**Palavras-chave:** Previsão de Evasão. Jogos Eletrônicos. Wavelets.

# Abstract

The online gaming market has rapidly grown in recent years, particularly since the rise of the service-based business model. As a result, the publishers of these games have started to share usual problems from the services business, like the profit erosion caused by customer churn. Predictive models have been used to address the churn problem in the mobile phones and credit cards markets, where companies have a huge volume of demographic and economic data about their customers. While game publishers often have only their users' email addresses.

The goal of this study is to propose a model for churn prediction based solely on the online games users' access patterns, where these time entries are fed into a set of operators that are able to analyze the data in the time-frequency plane domain, using the Discrete Wavelet Transform. Its main contribution is the input data parameterization proposed for the probabilistic classifiers based on the k-Nearest Neighbors algorithm.

Tested with real data from an online game users' access over a few months, the classifiers were evaluated using ROC (Receiver Operating Characteristic) and lift curves. The approach proposed in this thesis, based on the analysis of the time-frequency plane domain, has shown satisfactory results. Not only higher when compared with approaches based on both time or frequency domains, but also comparable to performances found on models with hundreds of predictive variables used in other markets.

**Keywords:** Churn Prediction. Games. Wavelets.

# Sumário

**Lista de Figuras**

**Lista de Tabelas**

**Lista de Símbolos**

**Lista de Abreviaturas**

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introdução</b>                                  | <b>1</b>  |
| 1.1      | Contexto . . . . .                                 | 1         |
| 1.2      | Motivação . . . . .                                | 4         |
| 1.2.1    | Vantagem Competitiva da Previsão . . . . .         | 6         |
| 1.2.2    | Impacto Econômico . . . . .                        | 8         |
| 1.2.3    | O Problema da Falta de Informações . . . . .       | 11        |
| 1.3      | Objetivos e Contribuições . . . . .                | 13        |
| 1.4      | Revisão Bibliográfica . . . . .                    | 14        |
| 1.5      | Organização da Tese . . . . .                      | 18        |
| <b>2</b> | <b>Conceitos Básicos</b>                           | <b>19</b> |
| 2.1      | Transformada Discreta de Wavelet . . . . .         | 19        |
| 2.1.1    | Transformada de Fourier de Curta Duração . . . . . | 19        |
| 2.1.2    | Transformada de Wavelet . . . . .                  | 21        |
| 2.1.3    | A Base Wavelet Haar . . . . .                      | 23        |
| 2.1.4    | Decomposição por Pacote de Wavelet . . . . .       | 23        |
| 2.1.5    | Análise de Frequência . . . . .                    | 26        |
| 2.1.6    | O Plano Tempo-Frequência . . . . .                 | 27        |



|          |   |           |
|----------|---|-----------|
| 2.2      | Avaliação de Classificadores . . . . .                    | 28        |
| 2.2.1    | Matriz de Confusão . . . . .                              | 29        |
| 2.2.2    | Curvas ROC . . . . .                                      | 32        |
| 2.2.3    | Área Abaixo da Curva ROC (AAC) . . . . .                  | 36        |
| 2.2.4    | Curvas de Elevação . . . . .                              | 37        |
| 2.2.5    | Elevação no Primeiro Decil (EPD) . . . . .                | 39        |
| <b>3</b> | <b>Metodologia</b>  | <b>41</b> |
| 3.1      | Definição do Problema . . . . .                           | 41        |
| 3.2      | Configurações de Testes . . . . .                         | 42        |
| 3.3      | Extração de Dados . . . . .                               | 44        |
| 3.4      | Pré-processamento . . . . .                               | 44        |
| 3.5      | Parametrizações da Entrada das Amostras . . . . .         | 45        |
| 3.5.1    | Proporção de Dias com Atividade (PDA) . . . . .           | 46        |
| 3.5.2    | Espectro de Potência (ESP) . . . . .                      | 47        |
| 3.5.3    | Plano Tempo-Frequência (PTF) . . . . .                    | 47        |
| 3.6      | Categorização das Amostras . . . . .                      | 48        |
| 3.7      | Função de Saída das Amostras . . . . .                    | 49        |
| 3.8      | Construção dos Conjuntos de Treinamento e Teste . . . . . | 49        |
| 3.9      | Classificação por <i>k-Nearest Neighbors</i> . . . . .    | 50        |
| 3.10     | Avaliação da Classificação . . . . .                      | 52        |
| 3.11     | Aplicação dos Classificadores . . . . .                   | 53        |
| <b>4</b> | <b>Resultados</b>   | <b>54</b> |
| 4.1      | Configuração PDA . . . . .                                | 54        |
| 4.2      | Configuração ESP16 . . . . .                              | 56        |
| 4.3      | Configuração ESP64 . . . . .                              | 58        |
| 4.4      | Configuração PTF16 . . . . .                              | 59        |
| 4.5      | Configuração PTF64 . . . . .                              | 61        |

|          |  |           |
|----------|--|-----------|
| 4.6      | Avaliação Comparativa . . . . .          | 62        |
| 4.7      | Discussão dos Resultados . . . . .       | 65        |
| <b>5</b> | <b>Conclusões e Considerações Finais</b> | <b>67</b> |
| 5.1      | Conclusões . . . . .                     | 67        |
| 5.2      | Trabalhos Futuros . . . . .              | 69        |
| 5.3      | Considerações Finais . . . . .           | 71        |
|          | <b>Referências</b>                       | <b>74</b> |

## 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:

