

Gestão de Projetos baseada em Aprendizado de Máquina

Igor Gustavo Hitzschky Lema*, Hélio Pedrini

Resumo

No contexto de engenharia de software, em conjunto com a tentativa de se otimizar a redução de custos, aumento de faturamento e produtividade, encontra-se o crescente volume de dados coletados no desenvolvimento de projetos. Uma atividade importante é entender a função de um gestor e a necessidade de uma metodologia que possibilite auxiliar a gestão de projetos. Este projeto de pesquisa investiga uma metodologia baseada em aprendizado de máquina para transformar esses dados em informações úteis.

Palavras-chave:

Aprendizado de Máquina, Gestão de Projetos, Análise de Dados.

Introdução

O desafio de melhorar a gestão de desenvolvimento de software pode ter diversas aplicações no mundo real. Cada vez mais, os dados impactam os negócios do dia a dia, assim se faz necessário desenvolver algoritmos capazes de analisar e gerar respostas em tempo hábil para este grande volume de dados.

Este projeto de pesquisa tem o objetivo de compreender o trabalho de gestores para que se possa classificar a qualidade de projetos e calcular o custo de desenvolvimento. Além disso, algoritmos de aprendizado de máquina foram investigados para uso na base de dados fornecida.

Resultados e Discussão

A base de dados é composta por indicadores mensurados manualmente durante o desenvolvimento de projetos de software. Para encontrar uma visão geral entre esses indicadores, gráficos de dispersão foram construídos e analisados exaustivamente.

Na construção de uma máquina de vetores de suporte (SVM), além de se fazer uso de métodos para calibrar parâmetros como o *GridSearch*, é necessário analisar os atributos com o objetivo de reduzir a dimensionalidade do problema.

Para selecionar os atributos de maior importância na construção de um classificador, foram aplicados quatro métodos presentes na literatura: *Recursive Feature Elimination*, *Random Forest Regressor*, *Support Vector Regression Weights* e *Univariate Feature Selection* (Figura 1).

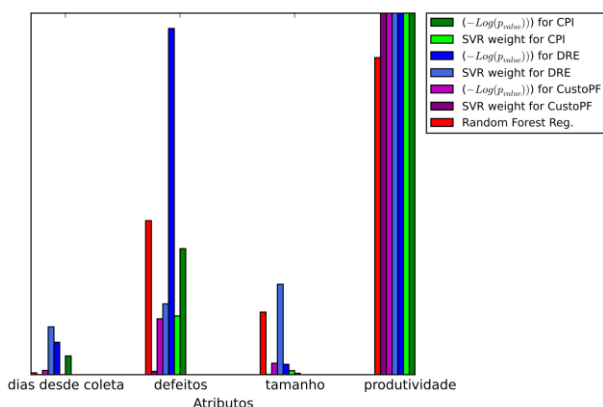


Figura 1. Resultados da avaliação de quatro métodos de seleção de atributos.

Separando os dados em treino e teste, pode-se treinar a máquina e, com o conjunto de testes, determinar a partir do algoritmo, o Coeficiente de Determinação e o Erro Absoluto Médio para verificar a acurácia da regressão.

O Coeficiente de Determinação, ou R^2 , é uma medida de ajuste de um modelo estatístico que varia entre 0 em 1, indicando, em porcentagem, o quanto o modelo consegue explicar os valores observados. Quanto maior o valor de R^2 , mais explicativo é o modelo, ou seja, melhor ele se ajusta à amostra de dados. O Erro Absoluto Médio é uma quantidade usada para mensurar quão próximo a predição é do eventual valor real.

Uma vez que a maioria dos indicadores é obtida manualmente, um fator importante em todas as etapas da metodologia é a confiabilidade do conjunto de dados.

Conclusões

A partir da análise dos resultados, pode-se concluir que, para a base de dados fornecida, é possível determinar com acurácia relativamente satisfatória o custo por ponto de função de um projeto de software em andamento.

Ainda que para outra base de dados aplicar a metodologia inteira pode verificar-se não prática, é necessário efetuar uma análise profunda sobre os dados.

Experimentos ainda se fazem necessários para verificar a aplicação do algoritmo em um contexto mais abrangente de desenvolvimento de projetos.

Agradecimentos

Agradeço ao meu orientador, prof. Dr. Hélio Pedrini, e ao prof. Dr. Luiz Fernando Bittencourt, por me auxiliarem durante todo o desenvolvimento do projeto de pesquisa. Agradeço, também, à empresa *Squadra Tecnologia S/A* por fornecer a base de dados do projeto. Este projeto foi financiado com recursos providos pela agência de fomento CNPq.

¹ R. P. Buse and T. Zimmermann. Analytics for software development. In Proceedings of the FSE/SDP Workshop on Future of Software Engineering Research, November 2010, ACM.

² J. Wen, S. Li, Z. Lin, Y. Hu, and C. Huang. Systematic literature review of machine learning based software development effort estimation models. *Information and Software Technology*, pp. 41-59, v. 54, n. 1, January 2012.