

Modelo de Transferência de Aprendizagem baseado em Regressão
Linear Regularizada

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Larissa Camila Ferreira da Silva (lcfs@cin.ufpe.br)
Orientador: Ricardo Bastos C. Prudêncio (nome@cin.ufpe.br)
Área: Aprendizagem de Máquina

Recife, 19 de Abril de 2017

Resumo

Muitos algoritmos da área de Aprendizagem de Máquina precisam de um bom conjunto de dados de treinamento para gerarem bons resultados nos dados de teste, onde estes dois tipos de dados devem ser de um mesmo contexto. Entretanto, em muitos problemas do mundo real não há dados suficientes para treinamento, e nestes casos se faz necessário o uso de técnicas de Transferência de Aprendizagem de um contexto para outro. Este trabalho apresenta uma técnica para modelagem de mudança de contexto em problemas de regressão linear. Neste método, a escolha de um atributo determina a divisão de um contexto em dois. Para que haja uma adaptação entre os ambientes, uma função de erro de aproximação é gerada através de uma regressão linear regularizada. Resultados experimentais com base em uma série de testes com dados reais serão mostrados.

Introdução

Cada vez mais, algoritmos e otimizações em aprendizagem de máquina têm se mostrado interessantes quando se buscam soluções para alguns problemas da área de inteligência artificial. Entretanto, diante de várias soluções existentes, a busca de uma solução ótima para um determinado problema é fundamental quando um sistema é pensado para, além de resolver uma tarefa específica, ter um desempenho satisfatório onde a sua computação não seja considerada demasiadamente custosa.

Um pressuposto dos algoritmos de aprendizagem de máquina diz que os dados de treinamento devem possuir o mesmo espaço de características e a mesma distribuição dos dados de teste, mas em muitas situações do mundo real esta hipótese não vem a se prender.

Portanto, em tais situações, o uso de transferência de aprendizagem se mostra uma solução de desempenho satisfatória, a qual pode ser usada tanto para problemas de regressão, classificação e agrupamento, quanto para problemas de mineração de dados.

Objetivos

O presente trabalho tem como objetivo a criação de um algoritmo para modelagem de mudança de contexto em problemas de regressão linear, feito através de técnicas de transferência de aprendizagem.

Como objetivos gerais, temos:

- Entender conceitos área de transferência de aprendizagem, partindo desde o grande tópico de aprendizagem de máquina, até a os subtipos de transferência comumente estudados e aplicados atualmente.
- Estudar as ferramentas computacionais apropriadas para o desenvolvimento de algoritmos na área de transferência de aprendizagem, no contexto desse trabalho.

Como objetivos específicos, temos:

- Propor e implementar um algoritmo que possa resolver um problema com dados reais utilizando conceitos de transferência de aprendizagem e ferramentas computacionais ágeis para o desenvolvimento.

Metodologia

1. Ferramentas para Desenvolvimento

Para a concretização da proposta desse trabalho de criação de um algoritmo para transferência de aprendizagem baseada em regressão linear regularizada, o conhecimento de como utilizar algumas tecnologias que envolvem programação numérica e científica e análise sobre os atributos de um conjunto de dados são fundamentais.

Abaixo serão mostradas tecnologias que podem auxiliar no desenvolvimento da aplicação proposta por este trabalho.

1.1 Python

Em um projeto de desenvolvimento que envolve programação científica e numérica é interessante a utilização de alguma ferramenta que permita com rapidez e corretude a solução de um problema, e para o contexto deste trabalho, a linguagem de programação Python foi escolhida pois é uma linguagem que permite um trabalho rápido e uma integração de sistemas com eficiência.

1.2 Biblioteca Scikit-learn

A linguagem Python em si já é muito poderosa para realização de cálculos matemáticos e experimentos científicos, mas para o contexto deste trabalho que é voltado para aprendizagem de máquina, se faz necessário a inclusão de uma biblioteca que pudesse facilitar o desenvolvimento dos modelos de aprendizagem.

A biblioteca SciKit-learn é uma biblioteca para Aprendizagem de Máquina para a linguagem Python, e a mesma é um projeto *open source* de esforço comunitário, ou seja, qualquer pessoa é aberta a contribuir. Ela possui ferramentas simples e eficientes para mineração de dados e análise de dados e é comercialmente usada em aplicações do *Spotify* e da *Evernote*^[1], por exemplo.

2. Realização de experimentos

Para testes nesse projeto, um conjunto de dados utilizado obtido no *UCI Machine Learning Repository*^[2] será utilizado, onde há uma quantidade expressiva de dados disponíveis para testes na área de Aprendizagem de Máquinas.

O conjunto de dados escolhido para realização de experimentos foi um referente a performance de estudantes de escolas públicas que moram na zona urbana e na zona rural. A tarefa de origem consiste em prever a nota de uma prova de Matemática ou Português de um aluno que mora na zona urbana e possui 307 instâncias, e a tarefa de destino prever a nota de um aluno que mora na zona rural possuindo 88 instâncias. Os atributos para ambas as tarefas são os mesmos e consistem de 30 atributos diferentes, que vão desde a idade do aluno, até a quantidade de horas estudada por ele por semana.

Cronograma

Referências

- [1] SCIKIT-LEARN: Machine Learning in Python. Machine Learning in Python. Disponível em: <<http://scikit-learn.org/stable/>>. Acesso em: 12 mar. 2017.
- [2] MACHINE Learning Repository: Center for Machine Learning and Intelligent Systems. Center for Machine Learning and Intelligent Systems. Disponível em: <<http://archive.ics.uci.edu/ml/>>. Acesso em: 12 mar. 2017.

Possíveis Avaliadores

Prof. Cleber Zanchettin (cz@cin.ufpe.br)

Prof. Germano Crispim Vasconcelos (gcv@cin.ufpe.br)

Assinaturas

Recife, ____ de _____ de _____

Larissa Camila Ferreira da Silva

(Aluno)

Ricardo Bastos C. Prudêncio

(Orientador)