

Mineração de dados

Mineração de fluxos de dados parcialmente classificados

link do vídeo: https://drive.google.com/file/d/1fQGA9u5Wbsqn3HB_BECP625IWq_laX/view

Aluno: Tiago Pastorello
Orientador: Fabiano Baldo

Mineração de dados

Por vezes chamado de "descoberta de conhecimento em bancos de dados", o termo "mineração" só foi cunhado nos anos 1990, e sua base compreende três disciplinas científicas: **estatística** (o estudo numérico das relações entre dados), **inteligência artificial** (inteligência exibida por softwares e/ou máquinas, que se assemelha à humana) e **machine learning** (algoritmos que podem aprender com dados). A tecnologia de mineração de dados continua evoluindo para acompanhar o potencial ilimitado do big data e a computação de baixo custo.

Mineração de fluxo de dados parcialmente classificados

Com mineração de fluxo é possível, melhorar o gerenciamento de dados pensando em velocidade, capacidade e escalabilidade; desenvolver técnicas de visualização de dados; encontrar novas oportunidades de negócio, etc.

Fluxo parcialmente classificado diz respeito a conjuntos de dados que são recebidos na forma de fluxo e apenas parte deles tem sua classe definida.

Mineração de fluxo de dados parcialmente classificados

Exemplo:

Um fluxo de dados de um sistema de telemetria de veículos, que traz as acelerações X, Y e Z mais a velocidade. Para alguns casos é sabido o tipo de condução do motorista, por exemplo (prudente, agressivo e etc). Entretanto, nem todos os dados recebidos tem a classe definida. Isso são fluxos parcialmente classificados.

Mineração de fluxo de dados parcialmente classificados

A aprendizagem de conceito pode depender de um contexto não presente no conjunto de dados inicial, ex: regras de previsão que variam dependendo da época do ano, previsões pluviométricas de acordo com as estações. Mudanças nesse contexto podem fazer com que ocorram mudanças no conceito a ser aprendido pelo indutor, fenômeno conhecido como mudança de conceito. Deste modo, um indutor criado para realizar aprendizagem de fluxos de dados com mudança de conceito, deve ser capaz de detectar a mudança de conceito, adequando o conceito aprendido, bem como deve ser robusto para não substituir indevidamente o conceito aprendido quando as instâncias forem ruidosas.

Mineração de fluxo de dados parcialmente classificados

Problema:

Como classificar fluxos de dados parcialmente classificados com presença de mudança de conceitos.

Pesquisa:

Aplicação de métodos de classificação semi-supervisionada baseada em árvore de decisão.

Mineração de dados

Mineração de fluxos de dados parcialmente classificados

Aluno: Tiago Pastorello
Orientador: Fabiano Baldo