

Identificação de Eventos Anormais em Cenas de Multidões

João Vitor Buscatto Silva*, Hélio Pedrini

Resumo

A compreensão do comportamento de multidões mostra-se um tópico de estudo importante nas áreas de segurança e vigilância. Neste contexto, vemos a necessidade de uma metodologia que possibilite a identificação de eventos anormais em tempo real de imagens de multidões. Este projeto de pesquisa investiga metodologias de identificação de anomalias já conhecidas da literatura e propõe um novo algoritmo, com base nos resultados encontrados.

Palavras-chave:

Anomalias, Imagens, Multidões

Introdução

As áreas de segurança e vigilância têm recebido grandes incentivos para se desenvolver, em que se destaca o estudo do comportamento coletivo, muito utilizado para a modelagem e compreensão do funcionamento de multidões.

O objetivo deste projeto de pesquisa é propor uma nova metodologia capaz de detectar eventos anormais em vídeos de multidões em tempo real. Espera-se que o algoritmo seja capaz de se adaptar a situações dinâmicas de anormalidade e possa classificar múltiplas anomalias em uma mesma cena. Bases de dados públicas são utilizadas para avaliar a metodologia proposta.

Resultados e Discussão

Uma das formas mais conhecidas pela literatura para efetuar a identificação de *outliers*, i.e. anomalias, é a aplicação de regressões no conjunto de dados estudado, resultando em uma curva denominada aqui como "curva ideal".

No algoritmo empregado, aplicamos três diferentes métodos de regressão: o *Linear Model Regression* (LMR), o *Support Vector Regression* (SVR) e a *Least Squares Regression* (LSR). Então, comparamos os resultados dessas regressões e escolhemos a que mais se adapta ao conjunto de dados estudado.

A partir dos valores das curvas ideais de cada regressão, podemos calcular o *Mean Squared Error* (MSE) da curva analisada e a *Cook's Distance* (CD), em que ambos são medidas de confiabilidade da regressão obtida.

A *Cook's Distance* de um elemento do vetor de regressão indica o nível de influência que este elemento exerce sobre o resto do conjunto. Ela é obtida comparando-se a regressão do conjunto original com uma regressão em que o elemento estudado é removido. Quanto maior a CD de um elemento, maior é sua influência sobre o conjunto de dados ao qual ele pertence. Com os vetores de CDs de cada vetor de regressão calculado, selecionamos o tipo de regressão que possui menor número de elementos com maior valor de CD, ou seja, a regressão que, em comparação com as demais, gera um vetor cujos elementos, individualmente, são menos influentes.

O MSE de uma regressão é dado pela média dos quadrados da diferença entre o vetor de dados original e o vetor dos dados da curva ideal, ou seja, ele mensura a

média da distância entre a curva ideal e o vetor de dados original. Quanto menor o MSE, mais adequada à base de dados é a regressão, sendo, por esse critério, a seleção da regressão pela métrica MSE.

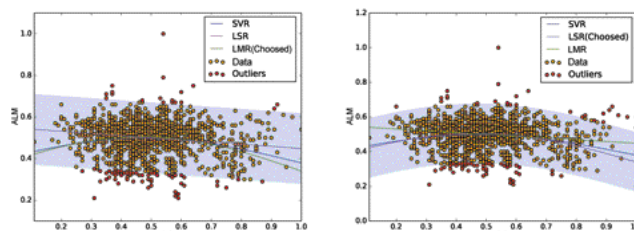


Figura 1. Gráficos das regressões LMR, SVR e LSR das variáveis 'MSG' (eixo X) e 'ALM' (eixo Y) da Yeast Data Set, da UCI[1]. A métrica MSE foi aplicada no exemplo mostrado à esquerda, escolhendo a LMR, e a métrica CD no exemplo à direita, escolhendo a LSR.

Conclusões

A partir da análise dos resultados aqui descritos, pode-se concluir que os métodos de detecção de *outliers* estudados têm uma convergência satisfatória no sentido da identificação dos *outliers*, já que frequentemente os mesmos dados são selecionados como anômalos por, pelo menos, dois dos três métodos em estudo. Entretanto, como já se era esperado, cada regressão estudada se comporta de forma diferente à base de dados em estudo, problema contornado pelo uso de métricas, como *Cook's Distance* e *Mean Squared Error*, que selecionam a mais adequada.

Os resultados obtidos até o momento têm se mostrado satisfatórios. Ainda é necessário efetuar testes adicionais, com mais bases de dados, e adaptar a metodologia desenvolvida para o contexto de imagens.

Agradecimentos

Agradeço ao meu orientador, prof. Dr. Hélio Pedrini, por me auxiliar durante todo o desenvolvimento do projeto de pesquisa. Este projeto foi financiado com recursos oriundos da agência de fomento CNPq.

[1] Yeast Data Set, UCI Machine Learning Repository. <https://archive.ics.uci.edu/ml/datasets/Yeast> - Acesso em 2016.