

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Aplicação de Reconhecimento de Padrões para Auxílio na Identificação de Criminalidade em Belo Horizonte

Igor Ferreira Bueloni¹

Centro Universitário de Belo Horizonte, UniBH, Belo Horizonte, MG

Wendell Ronald Mendes de Oliveira²

Centro Universitário de Belo Horizonte, UniBH, Belo Horizonte, MG

Flávio Henrique Batista de Souza³

Centro Universitário de Belo Horizonte, UniBH, Belo Horizonte, MG

1 Introdução

Este trabalho tem como propósito estudar a viabilidade de um complemento para ferramentas que utilizam a geolocalização de modo que se leve em consideração o índice de violência em uma região. Através de um estudo sobre a violência urbana, o modo como ela influencia uma cidade - principalmente os casos de turistas, que, em sua maioria, se tornam suas vítimas - e utilizando algoritmos de reconhecimento de padrões como Rede Neural Artificial (RNA) e Máquina de Suporte a Vetores (SVM), foi analisado o desempenho desses algoritmos juntamente com um conjunto de dados obtido no UCI Machine Learning Repository composto de 1994 amostras e 128 variáveis, como por exemplo: código do condado, código e nome da comunidade, etnia, idade, salário, número total de crimes violentos a cada 100 mil habitantes, dentre outras. Foram experimentadas seis variações da estrutura Multi Layer Perceptron (MLP) em treinamentos aplicando a técnica Set Validation. Na experimentação realizada com SVM, foram variadas as funções de kernel aplicando a técnica K-Fold Cross Validation com a finalidade de definir qual deles melhor se aplica a este conjunto de dados. Durante o processo de exploração, foi feita uma análise de cada atributo da base de dados e observados o comportamento e a relevância de cada um no resultado do algoritmo. Para comparação e análise entre esses algoritmos, bem como definir aquele que apresentou o melhor comportamento para a proposta principal do trabalho, foram utilizadas a curva ROC e AUC como avaliadores. [2]

2 Materiais e Métodos

Para os experimentos com a RNA MLP foram utilizadas 6 variações do algoritmo. Em cada tipo de RNA foram variados o número de neurônios na camada oculta em 1,

¹igorferreirasampaio@gmail.com

²wendell2309@gmail.com

³flavio.souza@prof.unibh.br

2, 5 e 10; e sua época em 1000, 2000, 3000 e 5000. O propósito destes experimentos foi gerar a AUC para analisar qual tipo de RNA, levando em consideração a quantidade de neurônios na camada oculta e época, apresenta o melhor resultado para o conjunto de dados analisado. Os tipos de RNA MLP Resilient Propagation e Backpropagation Weight Decay apresentaram os melhores valores de AUC, no caso, 0.9804 e 0.9765 respectivamente, dentre os 96 resultados.

Para os experimentos com a SVM foram utilizadas 9 opções de função kernel, e alteradas as configurações de custo e validação cruzada, respectivamente, nos valores 1, 2, 5, 10 e 3, 5, 10. Bessel Kernel foi a estrutura com melhor resultado, chegando a 0.969252. As funções Hyperbolic Tangent Kernel e Spline Kernel tiveram resultados baixos, mostrando-se inviáveis para este conjunto de dados. [1]

A estrutura que apresentou a melhor AUC em comparação com todos os resultados possui a seguinte estrutura: tipo de MLP Resilient Backpropagation, 5 neurônios na camada oculta e 3000 de época alcançando a AUC de 0.9804.

O terceiro experimento foi realizado retirando-se um a um os atributos do conjunto de dados (composto de 128 atributos no total) e comparando-se à AUC obtida (sem o atributo) com a AUC original (com o atributo), foi o que obteve o melhor resultado. Assim, pôde-se verificar o nível de ruído que cada atributo incidiu na rede, ora sendo de grande influência, melhorando a precisão da resposta; ora de menor influência, piorando a precisão da resposta. Foram encontrados 25 atributos com muita influência, 49 normais, 42 de pouca e 6 de nenhuma influência.

3 Conclusões

Como conclusões preliminares, é possível implementar um algoritmo em tecnologias de geolocalização baseado nos índices de criminalidade de uma região. A base de dados estudada apresentou atributos relevantes para os experimentos e na maioria dos testes obteve-se resultados com AUC acima de 0.9, o que a define como um conjunto de dados com baixo índice de ruídos entre os dados. Uma das contribuições deste trabalho é a demonstração via experimentos e testes exaustivos de variadas opções de algoritmos e técnicas de reamostragem para reconhecimento de padrões de violência que podem ser utilizados em aplicativos reais e com demandas atuais. Outra contribuição significativa é a análise de quais informações (campos) têm maior relevância e que devem ser coletadas para o preenchimento do conjunto de dados que será utilizado na aplicação (o complemento para navegadores).

Referências

- [1] D. Bonesso. Estimação dos Parâmetros do Kernel em um Classificador SVM na Classificação de Imagens Hiperespectrais em uma Abordagem Multiclasse, Dissertação de Mestrado em Sensoriamento Remoto, UFRGS, 2013.
- [2] A. P. Braga, A. P. L. F. C. Carvalho e T. B. Ludemir. *Redes Neurais Artificiais - Teoria e Aplicações*. LTC, Rio de Janeiro, 2011.