

Impacto da Quantidade de Características na Classificação do Multilayer Perceptron

Autor: Victor Hugo Souza Guimarães

E-mail: victor.guimaraes6@fatec.sp.gov.br

Orientador: Marcelo Buscioli Tenorio

RESUMO

Com a crescente área de *big data*, ciência de dados e aprendizado de máquina, principalmente aprendizagem profunda, diversas pesquisas estão sendo realizadas com o objetivo de obter a quantidade mínima necessária de características de um conjunto de dados. Para alcançar resultados computacionais satisfatórios (principalmente em relação ao tempo de processamento), estudos para reduzir a dimensionalidade dos dados são essenciais nos dias de hoje. O objetivo deste trabalho foi explorar a redução da quantidade de características e verificar a precisão na classificação de dados de imagem. A quantidade de características foi reduzida em 70% e o resultado da classificação manteve o percentual de verdadeiros positivos que o conjunto original possui.

Palavras-chave: Seleção de características, Classificação, Aprendizado de máquina.

ABSTRACT

With the growing area of big data, data science and machine learning, especially deep learning, several researches are being carried out with the aim of obtaining the minimum necessary number of characteristics from a data set. To achieve satisfactory computational results (mainly in relation to processing time), studies to reduce data dimensionality are essential nowadays. The objective of this work was to explore the reduction of the number of features and verify the accuracy in classifying image data. The number of features was reduced by 70% and the classification result maintained the percentage of true positives that the original set has.

Keywords: Features selection, Classification, Machine learning.

1 INTRODUÇÃO

A área de aprendizagem de máquina enfrenta dificuldade relacionada à dimensionalidade dos dados. A quantidade de características dos dados está diretamente relacionada com o tempo de resposta e resultados dos algoritmos (Silipo e Widmann, 2020). Sistemas de tempo real que reconhecem padrões, como face humana ou forma de fruta, precisam de tempo de resposta aceitável.

Especificamente alguns tipos de redes neurais artificiais, apresentam resultados insatisfatórios quando o conjunto de dados possui muitas características (Wójcik e Kurdziel, 2019). O ideal é sempre que possível, reduzir a quantidade de características e manter a acurácia no resultado.

O objetivo principal deste trabalho é explorar a redução da quantidade de características em conjuntos de dados e verificar como isso afeta a precisão na classificação de dados de imagem. Este artigo visa contribuir para os estudos na área de seleção de características e classificação de dados, com um foco especial no algoritmo seletor de características baseado em correlação e no algoritmo classificador *Multilayer Perceptron* (MLP). A seleção de características baseada em correlação identifica e seleciona um subconjunto de características que têm alta correlação com a variável de saída, mas são menos correlacionadas entre si. Já o *Multilayer Perceptron*, uma rede neural artificial de múltiplas camadas, é amplamente utilizado em tarefas de classificação devido à sua capacidade de modelar relações complexas entre características.

2 APRENDIZAGEM DE MÁQUINA

Aprendizagem de máquina é um conjunto de métodos que computadores usam para fazer e melhorar previsões ou comportamentos baseados nos dados (Alpaydin, 2020). A aprendizagem pode ser supervisionada, não supervisionada, semi-supervisionada ou por reforço. No aprendizado supervisionado é fornecido ao algoritmo de aprendizado, ou indutor, um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido.

Em geral, cada exemplo é descrito por um vetor de valores de características, ou atributos, e o rótulo da classe associada. O objetivo do algoritmo de indução é construir um

classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados, ou seja, exemplos que não tenham o rótulo da classe. Para rótulos de classe discretos, esse problema é conhecido como classificação e para valores contínuos como regressão. Já no aprendizado não-supervisionado, o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou clusters (Cheeseman e Stutz, 1990). Após a determinação dos agrupamentos, normalmente, é necessária uma análise para determinar o que cada agrupamento significa no contexto do problema que está sendo analisado.

É uma área de IA cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Onde um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas através da solução bem-sucedida de problemas anteriores que foram apresentando.

No aprendizado supervisionado hipóteses são induzidas utilizando um conjunto de exemplos de treinamento. Uma vez induzida, uma hipótese pode ter sua precisão avaliada utilizando-se os exemplos do conjunto de teste. Ela também pode ser avaliada de acordo com a sua consistência e completude. Já no aprendizado não supervisionado os padrões não são conhecidos a priori. Nesse caso, o objetivo é agrupar exemplos similares em clusters, de acordo com uma medida de similaridade pré-definida.

2.1 Seleção de Características

Uma área importante na aprendizagem de máquina é a seleção de características – algoritmos capazes de apresentar as características mais relevantes de um conjunto de dados. A seleção de características desempenha um papel crucial na construção de modelos de aprendizado de máquina eficazes, pois influencia diretamente a capacidade do modelo de generalizar a partir dos dados e realizar previsões precisas em novos exemplos não rotulados.

Em geral, cada exemplo é descrito por um vetor de valores de características, ou atributos, e o rótulo da classe associada. O objetivo do algoritmo de indução é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados, ou seja, exemplos que não tenham o rótulo da classe. Esses algoritmos reduzem a dimensionalidade dos dados, por vezes necessário.

Existem métodos dependentes e independentes do modelo. Os dependentes do modelo utilizam técnicas baseadas em redes neurais, neuro-fuzzy ou algoritmos genéticos, eles exploram as interações entre as características e a estrutura do modelo para determinar quais

características são mais relevantes para a tarefa em questão. Exemplos incluem técnicas baseadas em redes neurais, como a retropropagação com penalização de peso, métodos baseados em árvores de decisão, como o algoritmo CART (*Classification and Regression Trees*), e abordagens que utilizam algoritmos genéticos para otimizar conjuntos de características.

Os métodos independentes do modelo utilizam técnicas estatísticas (Campos, 2021). Esses métodos incorporam a seleção de características diretamente no processo de treinamento do modelo e avaliam as características sem levar em consideração a estrutura específica do modelo de aprendizado de máquina. Eles aplicam critérios estatísticos, como correlação, informação mútua ou testes de hipóteses, para medir a relevância de cada característica em relação ao rótulo da classe. Esses métodos podem ser mais rápidos e menos intensivos em computação do que os métodos dependentes do modelo, mas podem não capturar todas as interações complexas presentes nos dados.

2.2 Classificação

A classificação é uma aprendizagem de máquina supervisionada, como já foi mencionado, no aprendizado supervisionado o objetivo é induzir conceitos de exemplos que estão pré-classificados e estão rotulados com uma classe conhecida. De acordo com os valores atribuídos à classe, o problema é conhecido como classificação ou regressão. A classe - variável de saída independente - é predita a partir de características - variáveis de entrada dependentes (Alpaydin, 2020).

Classificação: Nesta abordagem, a variável de saída é categórica, o que significa que pertence a um conjunto finito de classes discretas. O objetivo é prever a categoria correta para novos exemplos com base em suas características. Exemplos típicos de problemas de classificação incluem a identificação de e-mails como "spam" ou "não spam", a classificação de imagens em diferentes categorias (por exemplo, "cachorro" vs. "gato"), e a previsão do tipo de doença com base em sintomas clínicos.

Regressão: Por outro lado, a regressão lida com variáveis de saída contínuas. Aqui, o objetivo é prever um valor numérico para novos exemplos com base em suas características. Exemplos de problemas de regressão incluem a previsão de preços de casas com base em suas características (tamanho, localização, número de quartos), a estimativa da demanda de energia elétrica, e a previsão de resultados financeiros, como lucros ou vendas futuras.

2.3 Validação Cruzada

A validação cruzada é uma técnica para avaliar a capacidade de generalização de um modelo, a partir de um conjunto de dados. O conceito central da técnica de validação cruzada é o particionamento do conjunto de dados em subconjuntos mutuamente exclusivos, e posteriormente, o uso de alguns destes subconjuntos para a estimação dos parâmetros do modelo (dados de treinamento) são os dados usados para treinar o modelo. Durante o treinamento, o algoritmo de aprendizado utiliza esses dados para ajustar os parâmetros do modelo de acordo com os padrões presentes nos dados. O objetivo é encontrar os parâmetros que minimizem a função de custo ou maximizem a função de pontuação, dependendo do tipo de problema (classificação, regressão, etc.). Os parâmetros ajustados do modelo são então utilizados para fazer previsões sobre novos dados, sendo os subconjuntos restantes (dados de validação ou de teste) empregados na validação do modelo (Kohavi, 1995, pp. 1137-1145). Após o treinamento do modelo, é essencial avaliar o desempenho do modelo em dados não utilizados durante o treinamento. Esta avaliação é realizada utilizando os dados de validação ou de teste. Se a validação cruzada for usada, esses dados serão chamados de conjunto de validação; se uma única divisão de treinamento/teste for usada, esses dados serão chamados de conjunto de teste. A ideia é verificar se o modelo generaliza bem para dados não vistos, ou seja, se ele é capaz de fazer previsões precisas em novos exemplos que não foram usados durante o treinamento.

3 TRABALHOS RELACIONADOS

Diversas pesquisas são realizadas na área de seleção de características, redução de dimensionalidade e classificação de dados. Um trabalho publicado analisa as técnicas de redução de dimensionalidade em big data (Reddy, 2020). Os resultados deste trabalho provam que os algoritmos de aprendizagem de máquina com a seleção de características por análise de componente principal, produzem melhores resultados quando a dimensionalidade dos conjuntos de dados é alta. Quando a dimensionalidade dos conjuntos de dados é baixa, observa-se que os algoritmos de aprendizagem de máquina sem redução de dimensionalidade produzem melhores resultados.

Por outro lado, o estudo também observa que em conjuntos de dados com baixa dimensionalidade, os algoritmos de aprendizado de máquina tendem a produzir melhores resultados sem a aplicação de técnicas de redução de dimensionalidade. Isso ocorre porque a aplicação desnecessária de redução de dimensionalidade pode remover informações

importantes e introduzir complexidade adicional no modelo, o que pode prejudicar seu desempenho.

4 METODOLOGIA

Para realização dos experimentos foi escolhido o software WEKA (Weka, 2019) e o conjunto de dados de frutas de um trabalho de graduação em computação (Nishida e Eler, 2014). Foram extraídas 52 características do conjunto de dados para descrever a imagem de frutas e 1 classe que identifica a fruta. São 163 imagens (instâncias) de 15 tipos de frutas. As características (numéricas) são: média, variância, desvio padrão das cores RGB e escala de cinza; coordenadas do contorno da imagem. O quadro 1 apresenta a quantidade de imagens em cada classe.

Quadro 1 - Quantidade de imagens em cada classe

Classe (Fruta)	Quantidade
1	8
2	7
3	16
4	15
5	9
6	6
7	11
8	13
9	11
10	14
11	10
12	7
13	13
14	11
15	12

A 1ª etapa - seleção das características - foi realizada com o algoritmo avaliador CfsSubsetEval (Hall, 1999) juntamente com o método de busca *BestFirst* e modo de validação cruzada.

Na 2ª etapa, as características não selecionadas na 1ª etapa foram excluídas do conjunto de dados original, gerando o primeiro subconjunto (22 características). Os demais subconjuntos

(de 21 até 2 características) foram obtidos excluindo as características selecionadas (da menor para a maior relevância) do resultado da 1ª etapa.

Foram gerados diversos subconjuntos de dados com quantidade reduzida de características. Todos os subconjuntos mantiveram a quantidade de instâncias (imagens) do conjunto original.

A 3ª etapa - classificação dos dados - foi realizada com o algoritmo classificador *Multilayer Perceptron* no conjunto de dados original e nos subconjuntos obtidos na 2ª etapa. A validação cruzada também foi utilizada nesta etapa.

A configuração padrão do software foi mantida nos algoritmos utilizados. Importante destacar que a validação cruzada utilizou 10 subconjuntos (folds), tanto para seleção de características quanto para classificação dos dados.

5 RESULTADOS E DISCUSSÕES

A seguir apresenta-se o resultado da classificação com o conjunto de dados original, composto por 52 características:

- Tempo para construir o modelo: 4,54 seg.
- Tempo para concluir a classificação: 43 seg.
- Instâncias classificadas corretamente: 79,14%
- Erro médio quadrático: 0,15
- Taxa de falso positivo: 0,01

O Quadro 2 apresenta as 22 características selecionadas na 1ª etapa, em ordem crescente da maior para a menor relevância. A coluna k-fold indica a quantidade de vezes (subconjuntos) que a característica foi selecionada.

Quadro 2 - Característica Selecionadas

Nº	k-fold	Característica
1	10	Média cor R
2	10	Média cor G
3	10	Variância cor R
4	10	Variância cor G
5	10	Coordenada x3
6	10	Coordenada x5
7	10	Coordenada y4
8	10	Coordenada y10
9	8	Média cor B

10	8	Desvio Padrão cor R
11	8	Desvio Padrão cor G
12	8	Coordenada y9
13	5	Coordenada y18
14	4	Coordenada x4
15	4	Coordenada y3
16	2	Média cor cinza
17	2	Variância cor cinza
18	2	Coordenada y7
19	2	Coordenada y12
20	1	Coordenada x7
21	1	Coordenada x11
22	1	Coordenada y5

O primeiro subconjunto composto por 22 características apresentou o seguinte resultado na classificação:

- Tempo para construir o modelo: 1,54 seg.
- Tempo para concluir a classificação: 16 seg.
- Instâncias classificadas corretamente: 78,52%
- Erro médio quadrático: 0,15
- Taxa de falso positivo: 0,01

O resultado (em média) da classificação dos subconjuntos obtidos na 2ª etapa, com 21, 20, 19, 18, 17, 16 e 15 características, foi:

- Tempo para construir o modelo: 1,11 seg.
- Tempo para concluir a classificação: 10 seg.
- Instâncias classificadas corretamente: 79,14%
- Erro médio quadrático: 0,15
- Taxa de falso positivo: 0,01

O resultado (em média) da classificação dos subconjuntos com 14, 13, 12, 11, 10, 9, 8, 7 e 6 características, foi:

- Tempo para construir o modelo: 0,79 seg.
- Tempo para concluir a classificação: 8 seg.
- Instâncias classificadas corretamente: 74,25%
- Erro médio quadrático: 0,16
- Taxa de falso positivo: 0,02

O Quadro 3 apresenta o resultado da classificação dos subconjuntos com 5, 4, 3 e 2 características.

Quadro 3 - Classificação dos Subconjuntos com 5, 4, 3 e 2 Características

Quantidade de Características	Instâncias Classificadas Corretamente (%)
5	68,71
4	65,03
3	60,73
2	34,96

6 CONSIDERAÇÕES FINAIS

A fase de seleção de características apresentou um bom resultado para geração do primeiro subconjunto de dados. Esse subconjunto com 42% da quantidade de características do conjunto original, apresenta um percentual de acerto de 78% na classificação, muito próximo do percentual de acerto (79%) do conjunto original.

Interessante observar que o subconjunto com apenas 28% da quantidade de características do conjunto original, ainda mantém o percentual de acerto de 79%. Este subconjunto com apenas 15 características tem a mesma representatividade nos dados que o conjunto original. Existem algumas características, que podem ser observadas no Quadro 2, determinantes para o processo classificatório.

Apesar da redução na quantidade de características, o algoritmo classificador *multilayer perceptron* apresenta resultados satisfatórios. É possível notar diminuição no tempo de processamento, um fator importante para a área de aprendizagem de máquina, principalmente quando são dados com alta dimensionalidade (muitas características) em sistemas de tempo real.

Os subconjuntos com 14, 13, 12, 11, 10, 9, 8, 7 e 6 características, ou seja, em média 19% da quantidade original de características, apresentam aproximadamente 74% de acerto na classificação. É um bom percentual de acerto, considerando a alta redução na quantidade de características. Diminuição de apenas 5% de acerto em relação ao conjunto original.

A partir de 5 características no subconjunto, o percentual de acerto na classificação cai bastante, entretanto, mantém-se na casa dos 60%.

Como trabalho futuro é relevante testar com outros algoritmos de seleção de características. Também é interessante explorar o significado das características selecionadas. Um outro trabalho futuro é realizar testes comparativos entre vários algoritmos de seleção, eliminando a fase de classificação.

REFERÊNCIAS

ALPAYDIN, E. Introduction to machine learning. MIT press, 2020.

CAMPOS, T. E. Técnicas de seleção de características com aplicações em reconhecimento de faces. Dissertação de Mestrado. São Paulo, Brasil, 2001.

CHEESEMAN, P.; STUTZ, J. Bayesian classification (autoclass): Theory and results advances in knowledge discovery and data mining, 1990.

HALL, M. A. Correlation-based Feature Subset Selection for Machine Learning. Tese de doutorado. Hamilton, Nova Zelândia, 1999.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: International Joint Conference on Artificial Intelligence, vol. XIV, 1995, pp. 1137–1145.

NISHIDA, M. M. H.; ELER, D. M. Estudo e implementação de algoritmos para reconhecimento de frutas. Trabalho de graduação. Presidente Prudente, Brasil, 2014.

REDDY, G. T. et al. Analysis of dimensionality reduction techniques on big data. IEEE Access, vol. VIII, 2020, pp. 54776-54788.

SILOPO, R.; WIDMANN, M. Three New Techniques for Data-Dimensionality Reduction in Machine Learning. Disponível em: <https://thenewstack.io/3-new-techniques-for-data-dimensionality-reduction-in-machine-learning/>. Acessado em 20 jul 2020.

WEKA. Waikato Environment for Knowledge Analysis. Versão 3.9.4. Universidade de Waikato, Hamilton, Nova Zelândia, 2019.

WÓJCIK, P. I.; KURDZIEL, M. Training neural networks on high-dimensional data using random projection. Pattern Anal Applic, v. 22, p. 1221–1231, 2019. DOI: 10.1007/s10044-018-0697-0.