

EEE950 - Redes Neurais Artificiais: Teoria e Aplicações

Exercício 1 - Histórico

José Geraldo Fernandes
Escola de Engenharia
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil

RESUMO

A proposta do artigo [1] é demonstrar matematicamente que a rede Perceptron de múltiplas camadas é um aproximador universal. Em outras palavras, demonstrar a capacidade de uma rede neural padrão com pelo menos uma camada escondida, neurônios suficientes, e função de ativação não linear, em aproximar qualquer função de interesse, Borel mensurável.

Para isso, o artigo formaliza definições para a arquitetura da rede, o mapeamento, e funções de ativação e com o teorema de Stone-Weirstrass, demonstra outros teoremas para alcançar a capacidade universal da rede. É importante considerar também que há uma preocupação de não se limitar a restrições espaciais estranhas na prática de aplicação desses modelos, como funções de ativação incomuns ou específicas e modificações exóticas na arquitetura como número de neurônios contínuo. A demonstração serve, portanto, para explicar o sucesso observado desse modelo em diferentes problemas práticos em vez de uma ocasionalidade do acaso.

Dada demonstração, o sucesso na aproximação da função não resta no poder do modelo mas de outros aspectos do problema de aprendizado. O operador pode, desse modo, preocupar-se com condições de treinamento, complexidade insuficiente ou não causalidade dos dados.

COMENTÁRIO

Uma grande frustração durante a pesquisa do modelo Perceptron foi sua incapacidade teórica de tratar problemas não linearmente separáveis. Com a implementação de múltiplas camadas e, atualmente, redes cada vez mais profundas e mais complexas, surge a preocupação sobre a limitação desses modelos. Afinal, no estudo de problemas difíceis, como processamento de linguagem natural e visão computacional, por exemplo, quer-se uma garantia que pelo menos a solução é possível, restando a atenção para problemas de infraestrutura como coleta e tratamento de dados e poder computacional.

O artigo responde essa demanda de maneira satisfatória, principalmente por abranger situações comuns na aplicação desses modelos, como generalização da função de ativação. Apesar disso, esse é apenas um primeiro passo na formulação matemática desses modelos. É observado, assim, um sucesso em diversos problemas do mundo real acompanhado por uma rasa intuição heurística de seu comportamento.

Outras demandas fora do escopo desse trabalho são duas: especificação de arquitetura; e, convergência do treinamento.

Uma pergunta óbvia que segue a leitura do artigo é dado um problema qual é o número "suficiente" de número de neurônios para um resultado aceitável. De forma mais geral, qual o desenho ótimo da rede dado um conjunto de dados. Essa pergunta não é respondida e é de grande importância já que essa determinação, na prática, depende da realização de diversos experimentos cobrindo as soluções de potencial, algo muitas vezes custoso. Ademais, qual a relevância da inclusão de camadas intermediárias uma vez que isso não é necessário na demonstração de universalidade. Um contraste grande com a aplicação, onde é observado sua essencialidade na solução dos problemas.

O segundo aspecto é qual a evolução esperada no treinamento. Por exemplo, seria interessante prever a convergência dado um problema, ou estimar tempo e recursos em um experimento. Até escolhas ótimas de hiper-parâmetros como taxa de aprendizado e tamanho de *batch* não necessitariam de buscas exaustivas.

Restam muitas demandas para um modelo pouco formalizado, apesar de relativamente recente, mas com grande sucesso em um conjunto amplo de problemas. Contudo, o artigo, de fato, serve como um primeiro passo nessa orientação cada vez mais valiosa.

REFERÊNCIAS

- [1] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.